UNIVERSITY
OF MICHIGAN
APR 21 1951
MATHEMATICS
LIBRARY

D

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

R. BAER
UNIVERSITY OF ILLINOIS

S. EILENBERG
COLUMBIA UNIVERSITY

D. C. LEWIS, JR.
THE JOHNS HOPKINS UNIVERSITY

A. WINTNER
THE JOHNS HOPKINS UNIVERSITY

WITH THE COÖPERATION OF

L. V. AHLFORS
C. B. ALLENDOERFER
W. L. CHOW
P. R. HALMOS
P. HARTMAN
I. KAPLANSKY
W. S. MASSEY
H. SAMELSON
J. J. STOKER
O. SZÁSZ
R. M. THRALL
H. WEYL

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXXIII, Number 2
APRIL, 1951

THE JOHNS HOPKINS PRESS
BALTIMORE 18, MARYLAND
U. S. A.

CONTENTS

PAGE

The theory of units of quadratic and hermitian forms. By K. G. RAMANATHAN, ... 233

The existence and limit behavior of the one-dimensional shock layer. By DAVID GILBARG, ... 256

Equivalences of measure spaces. By I. E. SEGAL, ... 275

Une extension du théorème de Vitali. Par ARNAUD DENJOY, . . 314

On normal operators in Hilbert space. By C. R. PUTNAM, . . . 357

Some retraction properties of the orbit decomposition spaces of periodic maps. By E. E. FLOYD, ... 363

On the non-existence of conjugate points. By AUREL WINTNER, . . 368

On the classical transcendents of mathematical physics. By PHILIP HARTMAN and AUREL WINTNER, ... 381

On the non-incerasing solutions of $y'' = f(x, y, y')$. By PHILIP HARTMAN and AUREL WINTNER, ... 390

The topology of regular curve families with multiple saddle points. By WILLIAM M. BOOTHBY, ... 405

On total nilpotent algebras. By ROY DUBISCH and SAM PERLIS, . 439

A theory of subinvariant Lie algebras. By EUGENE SCHENKMAN, . 453

Homomorphisms of a semigroup onto a group. By R. R. STOLL, . . 475

A note on a functional equation. By H. P. THIELMAN, . . . 482

The AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the JOURNAL remain on sale.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications should be sent to Professor AUREL WINTNER at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE 18, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

THE THEORY OF UNITS OF QUADRATIC AND HERMITIAN FORMS.*

By K. G. RAMANATHAN.

1. Introduction. In a recent paper Humbert [7] [1] carried through Siegel's reduction theory to indefinite quadratic or hermitian forms with coefficients which are integers of an algebraic number field K of finite degree over the field of rationals. He proved, among other theorems, the theorem that the number of classes of forms of given determinant, of order m, and with integral coefficients in K is finite.

It is the object of this paper to study the group of units of quadratic (hermitian) forms with integral elements in a finite algebraic extension K of the rationals. A unit of a form is a unimodular transformation leaving invariant the form. We prove

THEOREM A. *Let K be a totally real algebraic number field of finite degree over the field of rationals and $\mathfrak{S}$ the matrix of a quadratic form with elements integral in K. Let Γ be the group of units $\mathfrak{U}$ such that $\mathfrak{U}'\mathfrak{S}\mathfrak{U} = \mathfrak{S}$. Then* 1) *$\Gamma$ is finite if and only if $\mathfrak{S}$ is a totally definite symmetric matrix,* 2) *Γ is always finitely generated.*

The first part of Theorem A is proved in Section **4** and the second part in Section **5** since its proof requires the properties of fundamental domains.

A similar theorem is proved in Section **4** for hermitian forms. It is shown that the theory of hermitian forms works only in fields of the type $K_0 = K(d^{1/2})$ where K is a totally real field and d is a totally negative number of K.

We then generalise Siegel's results [8] on the measure of unit groups. We obtain for Γ/C where C is the centre of Γ, a faithful and discontinuous representation as a group B of non-Euclidean motions leaving invariant an open subspace J of a space of M dimensions, M being an integer depending on the signature of the quadratic form. In J we define a volume element dv invariant under B. Since B is discontonuous in J we can construct a fundamental domain F which is shown to have a finite number of neighbours.

* Received September 9, 1949; revised and substantially enlarged, March 30, 1950; slightly modified, October 12, 1950.

[1] Numbers in square brackets refer to bibliography at the end.

The important theorem stated as Theorem B asserts that $\int_F dv$ is finite except when $\mathfrak{S}$ is the matrix of a decomposable binary quadratic form. In this case it is shown that $\int_F dv$ is infinite. The proof of this depends on certain ideas used by Blumenthal [2] in the construction of a fundamental domain for the Hilbert modular group.

It is possible to prove an analogue of Theorem B for quadratic forms over number fields which are not necessarily totally real. These and other considerations regarding representations by indefinite quadratic forms will be shown elsewhere.

Finally I should like to record my thanks to Professor C. L. Siegel for his kindness in reading through the original version of the paper and making useful suggestions.

2. Notations. Symmetric or hermitian matrices will be denoted by capital German letters $\mathfrak{S}$, $\mathfrak{H} \cdots$ etc. $\mathfrak{U}$ will denote unimodular matrices, that is matrices whose elements are integers of K and whose determinant is an algebraic unit of K. $|\mathfrak{A}|$ will denote the determinant of a matrix $\mathfrak{A}$. abs $\mathfrak{A}$ will denote the absolute value of the determinant of $\mathfrak{A}$. We shall use abs a also to denote the absolute value of a number a. Small German letters $\mathfrak{x}, \cdots$ will be used as column vectors. $\mathfrak{S}[\mathfrak{U}]$ and $\mathfrak{S}\{\mathfrak{U}\}$ stand for $\mathfrak{U}'\mathfrak{S}\mathfrak{U}$ and $\mathfrak{U}'\mathfrak{S}\bar{\mathfrak{U}}$ where $\mathfrak{U}'$ and $\bar{\mathfrak{U}}$ are transposed and complex conjugate of $\mathfrak{U}$. $N(\mathfrak{T})$ denotes the product of $\mathfrak{T}$ and all its conjugates under the isomorphisms of K. $N(\mathfrak{T})$ will be the usual norm if $\mathfrak{T}$ is composed of elements of K. Similarly $\sigma(\mathfrak{T})$ is the spur of $\mathfrak{T}$. If $\mathfrak{S}$ is a matrix, we mean by the system $(\mathfrak{S})$, a matrix vector $(\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(f)})$. $\mathcal{E}_k$ will denote unit matrix of order k. $\mathfrak{O}$ will denote a zero matrix.

3. Reduction theory. Let K be an algebraic number field of finite degree $n = n_1 + 2n_2$ over the field of rationals and let $K^{(1)}, \cdots, K^{(n_1)}$ be the real conjugates of K and $K^{(n_1+\nu)}$, $K^{(n_1+n_2+\nu)}$, $\nu = 1, \cdots, n_2$ the pairs of complex conjugates of K. The unimodular matrices of order m in K form a multiplicative group Ω whose centre C consists of the matrices $\omega\mathcal{E}_m$ where ω is a root of unity in K. Let Γ be the factor group Ω/C. The problem of reduction consists in finding for Γ a fundamental domain in a suitable space. This problem was solved by Humbert [6] in the following way.

Let $\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(n_1)}$ be the matrices of n_1 non-singular positive definite quadratic forms in m variables and $\mathfrak{S}^{(n_1+\nu)}$, $\mathfrak{S}^{(n_1+n_2+\nu)}$; $(\nu = 1, \cdots, n_2)$ pairs of complex conjugate positive definite hermitian non-singular matrices of

order m. Let $\mathfrak{S}^{(i)}$, $(i = 1, \cdots, n_1 + n_2)$, be attached to the field $K^{(i)}$ and consider the system $(\mathfrak{S}) = (\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(n_1+n_2)})$. This system $(\mathfrak{S})$ defines a point in an open subspace P of a space of $\frac{1}{2}n_1 m(m+1) + n_2 m^2$ real dimensions. If $\mathfrak{U} \varepsilon \Gamma$, the mapping $(\mathfrak{S}) \to (\mathfrak{U}'\mathfrak{S}\bar{\mathfrak{U}})$ leaves the space P invariant. Since the algebraic units of K form a discrete subgroup of the continuous group in the space of $n_1 + n_2$ dimensions it follows that the representation $(\mathfrak{S}) \to (\mathfrak{S}\{\mathfrak{U}\})$ is discontinuous in P. Therefore we can construct a fundamental domain R in P for this discrete group of mappings. Let $\mathfrak{A}$ be a non-singular integral matrix in K and let $(\mathfrak{S}^*) = (\mathfrak{S}\{\mathfrak{A}\})$ be a system satisfying certain minimum conditions of Humbert. Then $\mathfrak{A}$ belongs to a finite set of matrices $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_\mu$ which are uniquely determined, up to multiplication on the left by a unimodular matrix $\mathfrak{U}$, by the field K and the integer m. For any system there is, in general, associated only one $\mathfrak{A}_i$. Since $\mathfrak{A} = \mathfrak{U}\mathfrak{A}_i$ for some $\mathfrak{A}_i$ and a unimodular $\mathfrak{U}$, we consider the system $(\mathfrak{S}^{**}) = \mathfrak{S}^*\{\mathfrak{A}_\nu^{-1}\}$. The totality of these points $(\mathfrak{S}^{**})$ constitute in P a *fundamental domain* R. The points $(\mathfrak{S}^*)$ define a point set R_0 called the *Humbert domain* defined in (35), (36), (37). Humbert showed that R is a union of μ convex pyramids whose faces have equations of the form $\tau(\mathfrak{S}\{\mathfrak{x}\}) = \tau(\mathfrak{S}\{\mathfrak{y}\})$, where $\mathfrak{x}$ and $\mathfrak{y}$ are independent column vectors in K. Putting $\mathfrak{x} = \mathfrak{a} + \mathfrak{b}$ and $\mathfrak{y} = \mathfrak{a} - \mathfrak{b}$ we write the above equation as

$$\tau(\mathfrak{a}'\mathfrak{S}\bar{\mathfrak{b}} + \bar{\mathfrak{a}}'\mathfrak{S}\mathfrak{b}) = 0, \tag{1}$$

where $\mathfrak{a} \neq \alpha\mathfrak{b}$ and α is a pure imaginary quantity.

In a recent paper [7] Humbert carried through the reduction theory to indefinite quadratic forms over algebraic number fields. Let us consider quadratic forms over totally real fields.[2] Let K be a totally real algebraic number field of finite degree f over the rationals and $K^{(1)} = K, \cdots, K^{(f)}$ its conjugates. Let $\mathfrak{S}^{(1)} = \mathfrak{S}$ be the matrix of a quadratic form in m variables and with integral elements in K. Let n_r, m_r $(= m - n_r)$ be the signature of $\mathfrak{S}^{(r)}$ where $\mathfrak{S}^{(r)}$ is the image of $\mathfrak{S}^{(1)}$ under the isomorphism $K^{(1)} \to K^{(r)}$, $(r = 1, \cdots, f)$. For each r there exists a real matrix $\mathfrak{C}^{(r)}$ so that

$$\mathfrak{S}^{(r)} = \mathfrak{C}^{(r)\prime} \begin{pmatrix} \mathcal{E}_{n_r} & \mathfrak{O} \\ \mathfrak{O} & -\mathcal{E}_{m_r} \end{pmatrix} \mathfrak{C}^{(r)}.$$

Following Hermite we associate with $\mathfrak{S}^{(r)}$ a positive definite matrix $\mathfrak{H}^{(r)} = \mathfrak{C}^{(r)\prime}\mathfrak{C}^{(r)}$ so that $|\mathfrak{H}^{(r)}| = \operatorname{abs} \mathfrak{S}^{(r)}$. Also $\mathfrak{H}^{(r)}$ satisfies

$$\mathfrak{H}^{(r)}\mathfrak{S}^{(r)-1}\mathfrak{H}^{(r)} = \mathfrak{S}^{(r)}, \qquad (r = 1, \cdots, f). \tag{2}$$

[2] The case of arbitrary number fields will be dealt with elsewhere.

We thus obtain a positive system $(\mathfrak{H})$ attached to $(\mathfrak{S})$. We consider all the positive systems $(\mathfrak{H})$ in the space P which satisfy the equations (2). They define a subspace H which is an irreducible algebraic manifold of

$$M = \sum_{r=1}^{f} n_r m_r \tag{3}$$

real dimensions. The points of H can be expressed in terms of M real parameters as follows.

$$\mathfrak{H}^{(r)} = 2\mathfrak{W}^{(r)} - \mathfrak{S}^{(r)}; \qquad \mathfrak{W}^{(r)} = \mathfrak{T}^{(r)-1}[\mathcal{E}\ \mathfrak{Y}^{(r)\prime}]; \tag{4}$$

$$\mathfrak{T}^{(r)} = \mathfrak{S}^{(r)-1}\begin{bmatrix}\mathcal{E}\\ \mathfrak{Y}^{(r)}\end{bmatrix} > 0, \qquad (r = 1, \cdots, f),$$

where $\mathfrak{Y}^{(r)}$ is a real matrix of n_r columns and m_r rows. (4) shows that the space

$$J: \ \mathfrak{T}^{(r)} > 0; \qquad (r = 1, \cdots, f) \tag{5}$$

of M real dimensions is mapped *upon* the space H of positive systems $(\mathfrak{H})$ satisfying (2). Let $\mathfrak{U}$ be a unimodular matrix in K such that $(\mathfrak{H}[\mathfrak{U}])$ is a point in the reduced space R; then we call $(\mathfrak{S}[\mathfrak{U}])$ *reduced*. Humbert showed [7] that for given value of $N(\mathfrak{S})$ the number of reduced integral systems $(\mathfrak{S})$ is finite.

In the case of hermitian forms a similar result can be proved. Let K_0 be an algebraic number field and $\mathfrak{R}$ its splitting field [1]. Then $\mathfrak{R}$ is a normal extension of the rationals. Let σ be an isomorphism of K_0 into its conjugate field K_1; then σ can be extended into an automorphism of $\mathfrak{R}$. Let τ be an automorphism of $\mathfrak{R}$ taking any quantity into its complex conjugate. A hermitian matrix with coefficients in K_0 is characterized by the property

$$a_{ij} = \bar{a}_{ji}, \tag{6}$$

so that for a hermitian matrix to go into a hermitian matrix by the isomorphism taking K_0 into its conjugate, it is necessary and sufficient that

$$\sigma\tau = \tau\sigma \tag{7}$$

for all automorphisms σ of $\mathfrak{R}$. Since τ is an element of order 2 we see that there is a subfield $\mathfrak{R}_0$ of $\mathfrak{R}$ which is invariant elementwise under $(1, \tau)$ where 1 is the identity automorphism. $\mathfrak{R}_0$ is a real subfield of $\mathfrak{R}$. If K_0 is real then by (7) all conjugates of K_0 would be real and our hermitian form would be merely a real quadratic form. Let us therefore take K_0 complex; then by (7) all its conjugates are complex. Let $\mathfrak{G} = K_0 \cap \mathfrak{R}$. Then $\mathfrak{G}$

is a real subfield of $\mathfrak{R}$. Also K_0 is of degree 2 over $\mathfrak{G}$. Let $\sigma_i K_0 = K_i$ for some automorphism σ_i; then K_i is of degree 2 over $\sigma_i\mathfrak{G}$ which is real. Thus $\mathfrak{G}$ is a totally real field and K_0 is an imaginary quadratic extension of $\mathfrak{G}$ such that all the conjugates of K_0 are complex. Thus $K_0 = K(d^{1/2})$ where K is totally real and d is a totally negative number of K. *It is therefore only in such fields that the theory of hermitian forms works.*

Let therefore K be a totally real field of degree $f \geq 1$ over the rationals and d a totally negative integer of K. Let $K^{(1)}(=K), \cdots, K^{(f)}$ be the conjugates of K. Let $K_0^{(r)} = K^{(r)}(d_r^{1/2})$ and $K_0^{(r+f)} = K^{(r)}(-d_r^{1/2})$ so that $\bar{K}_0^{(r)} = K_0^{(r+f)}$. Let $\mathfrak{S}^{(r)}$ be a hermitian matrix with integral elements in K_0 and of signature n_r, m_r. As before we find that the system $(\mathfrak{S}) = (\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(f)})$ has attached with it the totally positive system $(\mathfrak{H}) = (\mathfrak{H}^{(1)}, \cdots, \mathfrak{H}^{(f)})$, where

$$\mathfrak{H}^{(r)}\mathfrak{S}^{(r)-1}\mathfrak{H}^{(r)} = \mathfrak{S}^{(r)}, \qquad (r = 1, \cdots, f)). \tag{8}$$

These systems $(\mathfrak{H})$ constitute an algebraic manifold H_1 of $2M$ real dimensions and the points of H_1 can be parametrized by means of M complex quantities ($2M$ real quantities) $\mathfrak{Z}^{(1)}, \cdots, \mathfrak{Z}^{(f)}$ satisfying

$$\mathfrak{T}^{(r)} = \mathfrak{S}^{(r)-1}\begin{Bmatrix} \mathcal{E} \\ \mathfrak{Z}^{(r)} \end{Bmatrix} > 0, \qquad (r = 1, \cdots, f). \tag{9}$$

4. Group of units. Let K be a totally real field and $\mathfrak{S}$ the matrix of a quadratic form with integral coefficients. Let $\mathfrak{U}$ be a unimodular matrix such that

$$\mathfrak{U}'\mathfrak{S}\mathfrak{U} = \mathfrak{S}. \tag{10}$$

This implies that $\mathfrak{U}^{(r)\prime}\mathfrak{S}^{(r)}\mathfrak{U}^{(r)} = \mathfrak{S}^{(r)}$. Obviously $|\mathfrak{U}| = \pm 1$. We call $\mathfrak{U}$ *a unit of* $\mathfrak{S}$. There will be no confusion if we speak of the system $(\mathfrak{U}) = (\mathfrak{U}^{(1)}, \cdots, \mathfrak{U}^{(f)})$ as a unit of the system $(\mathfrak{S})$. The units $\mathfrak{U}$ obviously constitute a multiplicative group Γ_1 called the *group of units of* $\mathfrak{S}$.

Let us call a quadratic form $\mathfrak{x}'\mathfrak{S}\mathfrak{x}$ with integral elements in K totally definite if it and all its conjugates are definite forms. We shall now prove the

THEOREM 1. *Γ_1 is finite if and only if $\mathfrak{S}$ is the matrix of a totally definite form.*

To prove this theorem we require certain auxiliary lemmas.

LEMMA 1.[3] *A totally definite form* $\mathfrak{x}'\mathfrak{S}\mathfrak{x} = \sum_{i=1}^{m} a_i x_i^2$, *where* a_i *are integers of* K *has only a finite number of units.*

Proof. Let $\mathfrak{U}$ be a unimodular matrix so that $\mathfrak{U}'\mathfrak{S}\mathfrak{U} = \mathfrak{S}$. Let $\begin{pmatrix} u_{1k} \\ \vdots \\ u_{mk} \end{pmatrix}$ be the k-th column of $\mathfrak{U}$. Then since $\mathfrak{U}$ is a unit

$$\sum_{r=1}^{m} a_r u_{rk}^2 = a_k, \qquad (k = 1, \cdots, m). \tag{11}$$

Since the quadratic form is totally definite, a_r/a_s for all r and s are totally positive numbers of K. Thus

$$u_{rk}^2 \leqq a_k/a_r \leq \sigma[(a_1 + \cdots + a_m)(a_1^{-1} + \cdots + a_m^{-1})]. \tag{12}$$

Thus we get

$$\text{abs } u_{rk}^{(s)} \leq \{\sigma[(a_1 + \cdots + a_m)(a_1^{-1} + \cdots + a_m^{-1})]\}^{\frac{1}{2}} \tag{13}$$

for $s = 1, \cdots, f$. Thus by a well-known theorem of Kronecker the u_{rk} and hence $\mathfrak{U}$ have only a finite number of possibilities.

The Lemma follows at once if we observe that the group of all real linear transformations leaving a totally definite system $(\mathfrak{S})$ invariant is a compact Lie group and hence any discrete subgroup of it, like the group of units, must be finite.

LEMMA 2. *If* a *and* b *are integers of* K *and* ab *is not the square of an integer of* K *then the equation*

$$x^2 - aby^2 = 1 \tag{14}$$

has a finite or infinite number of solutions x, y *integral in* K *according as* ab *is or is not totally negative.*

This is a generalisation of the ordinary Pellian equation to algebraic number fields. The result is due to R. Fricke [4].

LEMMA 3. *Let* a *and* b *be integers of* K *and* ab *not a square. Then* $ax^2 - by^2$ *has an infinity of units unless* $(-ab)$ *is totally positive.*

Proof. If $-ab$ is totally positive then $ax^2 - by^2$ is totally definite and

[3] Here and elsewhere we write K for $K^{(\omega)}$ and $\mathfrak{S}$ for $\mathfrak{S}^{(\omega)}$. We use K in the theory of quadratic forms and K_0 in the theory of hermitian forms. When no superscripts appear, (11) etc. imply the corresponding relations got by taking the images in conjugate fields.

by Lemma 1 it has only a finite number of units. Let therefore $-ab$ be not totally positive. We now follow the method of Dirichlet-Dedekind [3].

Let $x \to \alpha x + \beta y;\ y \to \gamma x + \delta y,\ \alpha, \beta, \gamma, \delta$ integers of K, be a linear transformation leaving invariant the form $ax^2 - by^2$. Then

$$(15) \qquad a\alpha^2 - b\gamma^2 = a; \qquad a\beta^2 - b\delta^2 = -b; \qquad a\alpha\beta - b\gamma\delta = 0.$$

From (15) we deduce that

$$(15') \qquad \alpha = \delta = p, \qquad \gamma = aq, \qquad \beta = bq,$$

where p, q are integers of K satisfying

$$(16) \qquad p^2 - abq^2 = 1,$$

satisfy (15).[4] If ab is not a square in K, then from Lemma 2, there is an infinity of integral solutions of (16). Thus the number of solutions of (15) is infinite.

We can now prove Theorem 1. Let $\mathfrak{x}'\mathfrak{S}\mathfrak{x}$ be a non-degenerate quadratic form with integral elements in K. There exists a matrix $\mathfrak{C}$ in K (not necessarily integral) such that $\mathfrak{C}'\mathfrak{S}\mathfrak{C} = \mathfrak{D}$ is a diagonal matrix. If $\mathfrak{U}$ is a unit of $\mathfrak{S}$ then $\mathfrak{C}^{-1}\mathfrak{U}\mathfrak{C} = \mathfrak{V}$ satisfies $\mathfrak{V}'\mathfrak{D}\mathfrak{V} = \mathfrak{D}$. But $\mathfrak{V}$ is not in general integral. Let q be a rational integer such that $q\mathfrak{C}$ and $q\mathfrak{C}^{-1}$ are integral matrices. The matrices $\mathfrak{V} \equiv \mathcal{E} \pmod{q^2}$ form a subgroup of finite index in $\mathfrak{C}^{-1}\Gamma_1\mathfrak{C}$. The integral matrices $\mathfrak{V} \equiv \mathcal{E} \pmod{q^2}$ form a subgroup of finite index in the group of all units $\mathfrak{V}$. Thus the finiteness or otherwise of Γ_1 can be established from that of the units of the quadratic form $\mathfrak{x}'\mathfrak{D}\mathfrak{x}$. We may, as is readily seen, assume that the elements of $\mathfrak{D}$ are integers of K. Let therefore $\mathfrak{x}'\mathfrak{D}\mathfrak{x} = a_1x_1^2 + \cdots + a_mx_m^2$, $a_1, \cdots, a_m$ being integers of K. If $\mathfrak{x}'\mathfrak{D}\mathfrak{x}$ is totally definite then Lemma 1 shows that Γ_1 is finite. Let $\mathfrak{x}'\mathfrak{D}\mathfrak{x}$ be not totally definite. Then at least 2 of the quantities, say a_1 and a_2, are such that a_1a_2 is not totally positive. Let $\mathfrak{W}$ be a unimodular matrix of order $m-2$ of a linear transformation leaving $a_3x_3^2 + \cdots + a_mx_m^2$ invariant. The set of $\mathfrak{W}$ is not null since $\mathfrak{W} = \mathcal{E}$ is a possible value of $\mathfrak{W}$. Let $\mathfrak{V}$ be a unit of $a_1x_1^2 + a_2x_2^2$. Then

$$\mathfrak{U} = \begin{pmatrix} \mathfrak{V} & \mathfrak{O} \\ \mathfrak{O} & \mathfrak{W} \end{pmatrix}$$

is a unit of $\mathfrak{x}'\mathfrak{D}\mathfrak{x}$. Lemma 3 shows that if $-a_1a_2$ is not the square of an integer of K, there are an infinity of $\mathfrak{V}$. Suppose now that $-a_1a_2$ is the

[4] The solutions (15') are not all the solutions of (15). What we are concerned with is the finiteness or otherwise of the number of integral solutions and for this the crude results (15') are enough.

square of an integer of K. Then as before it suffices to show that $2x_1x_2$ has an infinity of units. Indeed if ϵ is an algebraic unit of K then

$$\mathfrak{V}_1 = \begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}, \qquad \mathfrak{V}_2 = \begin{pmatrix} 0 & \epsilon \\ \epsilon^{-1} & 0 \end{pmatrix}$$

leave $2x_1x_2$ invariant. Thus if $f > 1$ the number of units is infinite. If $f = 1$ it is trivial to see that $2x_1x_2$ has only a finite number of units. Theorem 1 is thus completely proved.

Exactly similar considerations hold good if we consider hermitian forms over algebraic number fields. We have only to modify Lemma 3.

LEMMA 4. *Let a and b be integers of K. Then $ax\bar{x} - by\bar{y}$ has an infinity of units in K_0 unless $-ab$ is totally positive in K.*

The method is similar to Lemma 3 and can be easily completed. The analogue of Theorem 1 for hermitian forms is

THEOREM 2. *A hermitian form $\mathfrak{x}'\mathfrak{S}\bar{\mathfrak{x}}$ with integral elements in K_0 has an infinity of units unless $\mathfrak{x}'\mathfrak{S}\bar{\mathfrak{x}}$ is totally definite.*

5. Fundamental domains. Let $\mathfrak{S}$ be the matrix of a non-degenerate quadratic form with integral elements in K. Let H be the space of positive systems $(\mathfrak{H})$ attached to $(\mathfrak{S})$. Let $\mathfrak{V}^{(r)}$ be a real non-singular matrix leaving $\mathfrak{S}^{(r)}$ invariant. Consider the system

$$(\mathfrak{V}) = (\mathfrak{V}^{(1)}, \cdots, \mathfrak{V}^{(f)}). \tag{17}$$

The systems $(\mathfrak{V})$ form a multiplicative group—multiplication being component wise: $(\mathfrak{V}_1\mathfrak{V}_2) = (\mathfrak{V}_1{}^{(1)}\mathfrak{V}_2{}^{(1)}, \cdots, \mathfrak{V}_1{}^{(f)}\mathfrak{V}_2{}^{(f)}) = (\mathfrak{V}_1) \cdot (\mathfrak{V}_2)$. Let Ω be this multiplicative group. In fact Ω is a Lie group. If $\mathfrak{U}$ is a unit of $\mathfrak{S}$ then $(\mathfrak{U}) = (\mathfrak{U}^{(1)}, \cdots, \mathfrak{U}^{(f)})$ is an element of Ω. Thus Γ_1, the group of units of $\mathfrak{S}$, is a discrete subgroup of Ω. It can easily be verified that if $(\mathfrak{H})$ is an element of H and $\mathfrak{V} \varepsilon \Omega$ then the system

$$(\mathfrak{V}'\mathfrak{H}\mathfrak{V}) = (\cdots, \mathfrak{V}^{(r)\prime}\mathfrak{H}^{(r)}\mathfrak{V}^{(r)}, \cdots)$$

is also in H. Thus Ω has a representation

$$(\mathfrak{H}) \to (\mathfrak{V}'\mathfrak{H}\mathfrak{V}) \tag{18}$$

as a group of homeomorphisms of H. If $\mathfrak{V} \varepsilon \Gamma_1$, then (18) gives a representation of Γ_1 in H. This representation is discontinuous in H since the group of unimodular matrices is discontinuous in P. (18) will be a faithful representation if we identify $\mathfrak{V}$ and $-\mathfrak{V}$, since both lead to the same point.

(18) thus gives a faithful and discontinuous representation of Γ where Γ is the group Γ_1 in which $\mathfrak{U}$ and $-\mathfrak{U}$ are identified. We now construct a fundamental domain for (18) in H.

From Humbert's theorem [7] on the finiteness of class number it is evident that in each class of the system $(\mathfrak{S})$ there is only a finite number of reduced systems. Let $(\mathfrak{S}[\mathfrak{U}_1]), \cdots, (\mathfrak{S}[\mathfrak{U}_g])$ be the reduced systems. Then $(\mathfrak{H}[\mathfrak{U}_1]), \cdots, (\mathfrak{H}[\mathfrak{U}_g])$ are reduced for certain $(\mathfrak{H})$ in H. Let $R_k = R(\mathfrak{U}_k^{-1})$ denote the set of points $(\cdots, \mathfrak{X}^{(r)}[\mathfrak{U}_k^{(r)-1}], \cdots)$ where $(\mathfrak{X}^{(1)}, \cdots, \mathfrak{X}^{(f)})$ is a point of R. Let $H \cap R_k = H_k$ and $H_1 + \cdots + H_g = G$. Then we can prove by exactly the same methods as in [8] that G does not contain in its interior (relative to H) two points equivalent under Γ. The same method shows that equivalent points on the boundary can only be finite in number. Thus G is bounded by a finite number of algebraic surfaces. G has only a finite number of neighbours $G_1, \cdots, G_k$ meaning thereby that if $G(\mathfrak{U})$ denote the image of G under the mapping $(\mathfrak{H}) \to (\mathfrak{U}'\mathfrak{H}\mathfrak{U})$, $\mathfrak{U} \varepsilon \Gamma$, then G and $G(\mathfrak{U})$ intersect in a non-null set only for $G(\mathfrak{U}) = G_1, \cdots, G_k$. Here we consider G as a closed set, closed relative to H. This would mean of course that if $\mathfrak{U}$ runs through all elements in Γ then $G(\mathfrak{U})$ gives a complete covering of H without gaps, there being overlapping only on the boundary. If in the above, $G_i = G(\mathfrak{U}_i)$, $\mathfrak{U}_i \varepsilon \Gamma$, then from the fact that H is a connected point set it follows that $\mathfrak{U}_1, \mathfrak{U}_2, \cdots, \mathfrak{U}_k$ form a set of generators of Γ. Hence the

THEOREM 1′. *The group Γ_1 of units of $\mathfrak{S}$ is finitely generated.*

Theorem A stated in Section **1** is thus completely proved.

In the case of hermitian forms also we may show that the group of units is finitely generated. To get a faithful representation of Γ_2, the group of units, we must identify $\mathfrak{U}$ and $\omega\mathfrak{U}$ where ω is a root of unity in K_0. (The roots of unity in K_0 form a finite group.)

6. Convergence of the volume integral. We shall now obtain a representation for Γ in the space of the parameters $y_{kl}^{(r)}$, $k = 1, \cdots, m_r$; $l = 1, \cdots, n_r$; $r = 1, \cdots, f$. We define the space J to be the open subspace

$$\mathfrak{X}^{(r)} = \mathfrak{S}^{(r)-1}\begin{bmatrix} \mathcal{E} \\ \mathfrak{Y}^{(r)} \end{bmatrix} > 0, \qquad (r = 1, \cdots, f), \tag{19}$$

of a space of M real dimensions. The transformation (4) maps the space J *upon* H. That the mapping (4) is upon may be deduced from the fact that (4) is a parametric representation of H or by following the method in [9].

Let

$$\mathfrak{V}'_r = \begin{pmatrix} \mathfrak{D}^{(r)} \mathfrak{C}^{(r)} \\ \mathfrak{B}^{(r)} \mathfrak{A}^{(r)} \end{pmatrix}$$

and $\mathfrak{V}'_r \mathfrak{S}^{(r)} \mathfrak{V}_r = \mathfrak{S}^{(r)}$, so that $(\mathfrak{V}) = (\mathfrak{V}_1, \cdots, \mathfrak{V}_f)$ is an element of Ω. Then it is easy to see that the transformation

$$(20) \qquad (\mathfrak{Y}) \to (\cdots, (\mathfrak{A}^{(r)} \mathfrak{Y}^{(r)} + \mathfrak{B}^{(r)})(\mathfrak{C}^{(r)} \mathfrak{Y}^{(r)} + \mathfrak{D}^{(r)})^{-1}, \cdots)$$

leaves the space J invariant. J is a space of M variables and (20) is a linear transformation of J into itself. (20) is thus a representation of Ω in J. If $\mathfrak{V} \varepsilon \Gamma$, then (20) gives a faithful and discontinuous representation of Γ in J. We can therefore, in J, construct a fundamental domain F bounded by a finite number of hyperplanes.

J is a symmetric riemannian space in the sense of E. Cartan. We can introduce easily a volume measure in this space. For let

$$(21) \qquad (\mathfrak{Y}_1) = (\cdots, (\mathfrak{A}^{(r)} \mathfrak{Y}^{(r)} + \mathfrak{B}^{(r)})(\mathfrak{C}^{(r)} \mathfrak{Y}^{(r)} + \mathfrak{D}^{(r)})^{-1}, \cdots);$$

$$\mathfrak{S}_1^{(r)} = \mathfrak{V}_r' \mathfrak{S}^{(r)} \mathfrak{V}_r, \qquad (r = 1, \cdots, f),$$

where $\mathfrak{V}_r$ is non-singular and $(\mathfrak{V}) = (\mathfrak{V}_1, \cdots, \mathfrak{V}_f)$ need not necessarily be in Ω. It is easy to see that the volume element

$$(22) \qquad dv = N \mid \mathfrak{S}^{(k)} \mid^{-nk/2} N \mid \mathfrak{T}^{(k)} \mid^{-m/2} d\mathfrak{Y}^{(1)} \cdots d\mathfrak{Y}^{(f)},$$

where

$$(23) \qquad d\mathfrak{Y}^{(r)} = \prod_{k=1}^{m_r} \prod_{l=1}^{n_r} dy_{kl}^{(r)},$$

so that $d\mathfrak{Y}^{(1)} \cdots d\mathfrak{Y}^{(f)}$ is the Euclidean volume element in J, remains invariant under (21). The volume element (22) has another interesting property. Of the pairs of integers n_r, m_r, $(r = 1, \cdots, f)$ there are some $n_r \leqq m_r$. Let us assume, without loss of generality, that

$$(24) \qquad n_r \leqq m_r, \quad (r = 1, \cdots, s); \qquad n_r > m_r, \quad (r = s+1, \cdots, f).$$

Let α be an integer in K such that

$$(25) \qquad \alpha^{(i)} > 0, \quad (i = 1, \cdots, s); \qquad \alpha^{(i)} < 0, \quad (i = s+1, \cdots, f).$$

That such integers α exist in K can easily be seen; for if $\omega_1, \cdots, \omega_f$ is a basis of numbers of K consider the inequalities

$$(26) \qquad \sum_{k=1}^{f} \beta_k \omega_k^{(i)} > 0, \quad (i = 1, \cdots, s); \qquad < 0, \quad (i = s+1, \cdots, f).$$

These f inequalities define f open half spaces in f dimensional Euclidean space. The intersection of these f half spaces is an open unbounded set and hence (26) has rational solutions $\beta_1, \cdots, \beta_f$.

Consider now the system $(\alpha\mathfrak{S})$ where α is an integer defined in (25). Then $(\mathfrak{S})$ and $(\alpha\mathfrak{S})$ have the same group of units. We shall show that the volume element (23) remains invariant if $(\mathfrak{S})$ is replaced by $(\alpha\mathfrak{S})$ and at the same time we interchange n_r, m_r, $(r = s+1, \cdots, f)$. Now let $\mathfrak{B}_r = (\alpha^{(r)})^{\frac{1}{2}}\mathcal{E}$, $(r = 1, \cdots, s)$ and $\mathfrak{B}_r = (-\alpha^{(r)})^{\frac{1}{2}}\mathcal{E}$, $(r = s+1, \cdots, f)$. Then by (21) the volume element remains unchanged. We have therefore only to consider invariance under changing $\mathfrak{S}^{(r)}$ into $-\mathfrak{S}^{(r)}$, $(r = s+1, \cdots, f)$, and simultaneously interchanging n_r, m_r, $(r = s+1, \cdots, f)$. This is but an immediate consequence of a corresponding theorem of Siegel's [8].

Let us call a binary quadratic form (hermitian form) *decomposable* if by a linear transformation in K (in K_0) its matrix can be brought to the form

$$\begin{pmatrix} 0 & a \\ a & b \end{pmatrix}; \qquad \begin{pmatrix} 0 & \bar{a} \\ a & b \end{pmatrix}, b = \bar{b}.$$

We shall now prove the

THEOREM B. *Let $\mathfrak{S}$ be not the matrix of a decomposable binary quadratic form; then the volume of a fundamental domain F in J is finite. In other words*

$$N \mid \mathfrak{S}^{(k)} \mid^{-n_k/2} \int_F N \mid \mathfrak{T}^{(k)} \mid^{-m/2} d\mathfrak{Y}^{(1)} \cdots d\mathfrak{Y}^{(f)} \tag{27}$$

converges.

We have made an exception of binary decomposable forms. In Section 7 we shall show that in this case the integral (27) diverges.

The integral in (27) gives a measure of the group of units, Γ. This measure is the same for two isomorphic groups. Since the units of $\mathfrak{S}$ and $\alpha\mathfrak{S}$ are isomorphic, their unit groups have the same measure. We shall therefore, in our proof, assume that the system of signatures of $\mathfrak{S}$ itself satisfies

$$n_r \leq m_r, \qquad (r = 1, \cdots, f). \tag{28}$$

Furthermore, since two matrices in a class have their unit groups isomorphic, it follows that (27) depends only on the class of the system $(\mathfrak{S})$. Theorem B would have followed very simply from the properties of measures provided F was compact (relative to J). But this is not in general the case.

We formulate the corresponding theorem in the case of hermitian forms. It is easy to see that the invariant volume element is [5]

---

[5] The volume element (29) can be computed thus. Let $\mathfrak{Z}_1 = (\mathfrak{A}\mathfrak{Z} + \mathfrak{B})(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$. Let $\delta\mathfrak{Z}$ denote the matrix of differentials $\delta x_{kl} + i\delta y_{kl}$. Then $(\delta z'_{kl}) = \delta\mathfrak{Z}_1 = (\mathfrak{A} - \mathfrak{Z}_1\mathfrak{C})$

$$(29) \qquad dv = 2^{M_1} N \,|\, \mathfrak{S}^{(k)} \,|^{-n_k/2 N} \,|\, \mathfrak{T}^{(k)} \,|^{-m/2}\, d\mathfrak{Z}^{(1)} \cdots d\mathfrak{Z}^{(f)},$$

where $\mathfrak{Z}^{(r)}$ is a complex matrix of m_r rows and n_r columns. Thus

$$(30) \quad \mathfrak{Z}^{(r)} = \mathfrak{X}^{(r)} + i\mathfrak{Y}^{(r)}; \quad \mathfrak{X}^{(r)} = (x_{kl}{}^{(r)}), \quad \mathfrak{Y}^{(r)} = (y_{kl}{}^{(r)}), \quad i^2 = -1;$$

$$d\mathfrak{Z}^{(r)} = \prod_{k,l} dx_{kl}{}^{(r)} dy_{kl}{}^{(r)}.$$

THEOREM B′. *If F is a fundamental domain for the transformations*

$$(\mathfrak{Z}) \to (\cdots, (\mathfrak{A}^{(r)}\mathfrak{Z}^{(r)} + \mathfrak{B}^{(r)})(\mathfrak{C}^{(r)}\mathfrak{Z}^{(r)} + \mathfrak{D}^{(r)})^{-1}, \cdots)$$

where $\begin{pmatrix} \mathfrak{D}^{(r)}, \mathfrak{B}^{(r)} \\ \mathfrak{C}^{(r)}, \mathfrak{A}^{(r)} \end{pmatrix}'$ *is a unit of the hermitian form* $\mathfrak{x}'\mathfrak{S}\bar{\mathfrak{x}}$ *with integral elements in* K_0 *in the subspace*

$$(31) \qquad (\mathcal{E}\mathfrak{Z}^{(r)\prime})\mathfrak{S}^{(r)-1}\left(\frac{\mathcal{E}}{\mathfrak{Z}^{(r)}}\right) > 0, \qquad (r = 1, \cdots, f),$$

of the space of $M_1 = 2M$ *real variables, then*

$$(32) \qquad \int_F dv < \infty,$$

where dv is the volume element (29).

It must be noticed that unlike binary decomposable quadratic forms, the units of binary decomposable hermitian forms do *not* have infinite measure.

We shall now prove Theorems B and B′. We shall give in detail the proof of Theorem B. That of B′ is quite similar and we make a few remarks about it at the end of this section.

As pointed out before we shall assume that the system of signatures of $(\mathfrak{S})$ satisfies (28). The transformations (4) show that to every point in J corresponds only one point in H. Let $\mathfrak{U}_1, \cdots, \mathfrak{U}_g$ be the unimodular matrices such that $(\mathfrak{S}[\mathfrak{U}_1]), \cdots, (\mathfrak{S}[\mathfrak{U}_g])$ are reduced. Then for some $\mathfrak{H}$ $(\mathfrak{H}[\mathfrak{U}_1])$, $\cdots$, $(\mathfrak{H}[\mathfrak{U}_g])$ are reduced in the sense of Humbert. F is the image in J of a fundamental domain G in H. $G = H_1 + H_2 + \cdots + H_g$, where H_a denotes the set of $(\mathfrak{H})$ in G for which $(\mathfrak{H}[\mathfrak{U}_a])$ is reduced. Also $(\mathfrak{H}[\mathfrak{U}_a])$ satisfies

$$(\mathfrak{H}_a) = (\mathfrak{H}[\mathfrak{U}_a]), \qquad (\mathfrak{S}_a) = (\mathfrak{S}[\mathfrak{U}_a]), \qquad \mathfrak{H}_a{}^{(r)}\mathfrak{S}_a{}^{(r)-1}\mathfrak{H}_a{}^{(r)} = \mathfrak{S}_a{}^{(r)}.$$

$\times \delta\mathfrak{Z}(\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$. Let $\mathfrak{U} = (u_{kl}) = (\mathfrak{A} - \mathfrak{Z}_1\mathfrak{C})$, $\mathfrak{V} = (v_{kl}) = (\mathfrak{C}\mathfrak{Z} + \mathfrak{D})^{-1}$. Then $\delta z'_{kl} = \sum_{i,p} u_{ki}\delta z_{ip} v_{pl}$. The Jacobian of the transformation is thus a matrix of order $2mn$. Let $\left|\frac{d\mathfrak{Z}_1}{d\mathfrak{Z}}\right|$ denote the absolute value of this Jacobian. Then

$$\left|\frac{d\mathfrak{Z}_1}{d\mathfrak{Z}}\right| = \begin{vmatrix} \mathfrak{B} & \mathfrak{T} \\ -\mathfrak{T} & \mathfrak{B} \end{vmatrix} = \operatorname{abs}(\mathfrak{B} + i\mathfrak{T})^2,$$

where $\mathfrak{B} = (Ru_{ki}Rv_{pl} - Iu_{ki}Iv_{pl})$, $\mathfrak{T} = (Iu_{ki}Rv_{pl} + Ru_{ki}Iv_{pl})$ and R and I denote real and imaginary parts.

Let F_a denote that part of F which corresponds to H_a. Then

$$\int_F dv = \sum_{a=1}^{g} \int_{F_a} dv. \tag{33}$$

With every positive system $(\mathfrak{H})$ is associated a unique integral matrix belonging to the set $\mathfrak{A}_1, \cdots, \mathfrak{A}_\mu$. Let $\mathfrak{A}_t$ correspond to the reduced system $(\mathfrak{H}_a)$. Then $(\mathfrak{H}_a[\mathfrak{A}_t])$ is a point in the Humbert space R_0. We subdivide F_a into $F_{a1}, \cdots, F_{at}$, where F_{at} is that part of F_a for which the corresponding $(\mathfrak{H}_a)$ which are reduced are attached to $\mathfrak{A}_t$. Using the invariance of the volume element under (21) we get

$$\int_{F_a} dv = \sum_{t=1}^{\tau} \int_{F_{at}} dv. \tag{34}$$

The $(\mathfrak{H}_a[\mathfrak{A}_t])$ are points of the Humbert space R_0. Also $(\mathfrak{S}_a[\mathfrak{A}_t])$ are integral matrices and if we write $(\mathfrak{H}) = (\mathfrak{H}_a[\mathfrak{A}_t])$, $(\mathfrak{S}) = (\mathfrak{S}_a[\mathfrak{A}_t])$, then $\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H} = \mathfrak{S}$. To prove Theorem B it is necessary to prove $\int_{F_{at}} dv < \infty$ for all a and t. Consider the points $(\mathfrak{H})$ satisfying

$$0 < h_k^{(\lambda)} \leq c_1 h_l^{(\mu)}, \qquad k \leq l, \tag{35}$$

$$\operatorname{abs} h_{kl}^{(\lambda)} \leq c_2 h_k^{(\lambda)}, \qquad k \leq l, \tag{36}$$

$$h_1^{(\lambda)} \cdots h_m^{(\lambda)} \leq c_3 \mid \mathfrak{H}^{(\lambda)} \mid, \tag{37}$$

where $\mathfrak{H}^{(\lambda)} = (h_{kl}^{(\lambda)})$, and $h_k^{(\lambda)} = h_{kk}^{(\lambda)}$, $(k = 1, \cdots, m)$; c_1, c_2, c_3 are constants depending only on K and m. This point set $R_0(c_1, c_2, c_3)$ is a closed point set called the *Humbert space* $R_0(c_1, c_2, c_3)$, and contains all the points $(\mathfrak{H}[\mathfrak{A}])$ where $(\mathfrak{H})$ is a point in the reduced domain R and $\mathfrak{A}$ is a suitable one of the set $\mathfrak{A}_1, \cdots, \mathfrak{A}_\mu$. Let c_1, c_2, c_3 be chosen large enough, in particular $c_1 > 1$. Consider the point set X_{at} in J such that the points $(\mathfrak{H}_a[\mathfrak{A}_t])$ corresponding to the points in X_{at} lie in $R_0(c_1, c_2, c_3)$. Obviously F_{at} is contained in X_{at}. For the proof of Theorem B it suffices to show that $\int_{X_{at}} dv < \infty$. Because of the invariance of the volume element under (21) it suffices to prove that $\int dv$ for those points in the J space for which the corresponding $(\mathfrak{H})$ satisfy (35), (36), (37) is finite. The success of the proof depends largely on the fact that $c_1 > 1$ and on the inequality (35) which does not depend on λ or μ.

Since $\mathfrak{H}^{(r)}$ is positive definite, $h_1^{(r)}, \cdots, h_m^{(r)}$ are all positive and it can be easily proved that (if $\mathfrak{S}^{(r)} = (s_{kl}^{(r)})$)

$$\operatorname{abs} s_{kl}^{(r)} \leq (h_k^{(r)} h_l^{(r)})^{\frac{1}{2}}, \qquad (r = 1, \cdots, f). \tag{38}$$

Since $|\mathfrak{S}^{(r)}| \neq 0$, at least one term in its expansion is not zero. Also $\mathfrak{S}^{(r)}$ is an integral matrix so that it is possible to find for each k an $l = l_k$ such that

$$1 \leq N(s_{kl_k}{}^{(r)}) \leq N(h_k{}^{(r)} h_{l_k}{}^{(r)})^{\frac{1}{2}}. \tag{39}$$

Since $N(h_k{}^{(r)} h_{l_k}{}^{(r)}) \geq 1$, there is a μ depending on k and r such that

$$h_k{}^{(\mu)} h_{l_k}{}^{(\mu)} \geq c_1{}^{-2}, \tag{40}$$

for otherwise $N(h_k{}^{(r)} h_{l_k}{}^{(r)}) < c_1{}^{-2f} < 1$, which is false.

Let us fix this μ and let a run over the integers $1, 2, \cdots, m$. Then of the a numbers $l_1, l_2, \cdots, l_a$ at least one of them, say l_k is $\leq m - a + 1$. Hence

$$k \leq a; \qquad l_k \leq m - a + 1. \tag{41}$$

Using (35) with the μ in (40) we get

$$h_a{}^{(r)} \geq c_1{}^{-1} h_k{}^{(\mu)}; \qquad h_{m-a+1}{}^{(r)} \geq c_1{}^{-1} h_{l_k}{}^{(\mu)}. \tag{42}$$

From (40) and (42) follows

$$h_a{}^{(r)} h_{m-a+1}{}^{(r)} \geq c_1{}^{-4}, \qquad (r = 1, \cdots, f). \tag{43}$$

This inequality is true for all $a = 1, \cdots, m$.

Consider now the products $h_k{}^{(r)} h_{m-k}{}^{(r)}$; $(k = 1, \cdots, m-1)$. There is an integer q_r $(0 \leq q_r \leq [m/2])$ such that

$$h_k{}^{(r)} h_{m-k}{}^{(r)} \geq c_1{}^{-4}, \qquad (q_r < k < m - q_r); \tag{44}$$

$$h_{q_r}{}^{(r)} h_{m-q_r}{}^{(r)} < c_1{}^{-4}. \tag{45}$$

q_r might be zero or $[m/2]$. In the first case we do not consider (45) and in the second case we ignore (44). It should be noted that the integers $q_1, \cdots, q_f$ depend only on the point in the J space.

Let $k \leq q_r$ and $l \leq m - q_r$. Then using (35) and (45) we immediately have

$$h_k{}^{(\lambda)} h_l{}^{(\lambda)} < c_1{}^{-2} \qquad \text{(for all } \lambda), \tag{46}$$

and using (38) we obtain

$$0 \leq N(\text{abs } s_{kl}{}^{(r)}) < c_1{}^{-2f} < 1, \tag{47}$$

which means $s_{kl}{}^{(r)} = 0$. Thus

$$s_{kl}{}^{(r)} = 0 \qquad \text{for all } r,\ k \leq q_r,\ l \leq m - q_r. \tag{48}$$

(48) shows that if $q = \text{Max}(q_1, \cdots, q_f)$, then for all λ, $\mathfrak{S}^{(\lambda)}$ has a decomposition

$$\mathfrak{S}^{(\lambda)} = \begin{pmatrix} \mathfrak{O}, & \mathfrak{O}, & \tilde{\mathfrak{P}}_\lambda' \\ \mathfrak{O}, & \tilde{\mathfrak{N}}_\lambda, & \tilde{\mathfrak{Q}}_\lambda' \\ \tilde{\mathfrak{P}}_\lambda, & \tilde{\mathfrak{Q}}_\lambda, & \tilde{\mathfrak{G}}_\lambda \end{pmatrix}, \tag{49}$$

where $\tilde{\mathfrak{N}}_\lambda$ is $(m-2q)$-rowed symmetric and $\mathfrak{G}_\lambda$ is q-rowed symmetric. We may, without contradiction, restrict ourselves to $\mathfrak{S}^{(r)}$ and write

$$\mathfrak{S}^{(r)} = \begin{pmatrix} \mathfrak{O} & \mathfrak{O} & \mathfrak{P}_r' \\ \mathfrak{O} & \mathfrak{N}_r & \mathfrak{Q}_r' \\ \mathfrak{P}_r & \mathfrak{Q}_r & \mathfrak{G}_r \end{pmatrix}, \tag{50}$$

where $\mathfrak{N}_r$ is $m-2q_r$-rowed symmetric and $\mathfrak{G}_r$ is q_r-rowed symmetric respectively.

The essential point of the inequalities (44) and (45) is to obtain:

$$\begin{aligned} &h_1^{(r)}, \cdots, h_{q_r}^{(r)} && \text{are bounded;} \\ &h_{q_r+1}^{(r)}, \cdots, h_{m-q_r}^{(r)} && \text{are bounded together with their reciprocals;} \\ &h_{m-q_r+1}^{(r)}, \cdots, h_m^{(r)} && \text{have their reciprocals bounded,} \end{aligned} \tag{51}$$

these bounds depending on $|\mathfrak{H}|$, K and m only. To prove them we observe the identity

$$\prod_{k=1}^{m} (h_k^{(r)} h_{m-k+1}^{(r)}) = \prod_{k=1}^{q_r} (\cdots)^2 \prod_{k=q_r+1}^{m} (\cdots) / \prod_{k=1}^{q_r} (\cdots). \tag{52}$$

Now $q_r \leq m - q_r$ so that

$$\prod_{k=1}^{m} (\cdots) \geq h_{m-q_r}^{(r)2} \prod_{k=1}^{q_r} (\cdots)^2. \tag{53}$$

From this we obtain, using (37) and (43), that

$$h_{m-q_r}^{(r)} \leq c_4, \tag{54}$$

c_4 being a constant depending on $|\mathfrak{H}|$, K and m. Now using (44) we have

$$(h_{m-k}^{(r)})^{-1} \leq c_5 h_k^{(r)}, \qquad (q_r < k < m - q_r). \tag{55}$$

Therefore from (54)

$$(h_{m-k}^{(r)})^{-1} \leq c_6, \qquad (q_r < k < m - q_r). \tag{56}$$

Inequalities (54) and (56) together prove (51).

From now on the proof is as in Siegel's paper [8]. We write down only the essential results which may be proved easily.

$$|\mathfrak{X}^{(r)}|^{-m/2} \leq c_7 \prod_{k=1}^{q_r} (h_k^{(r)})^{m/2}, \qquad q_r \leq n_r. \tag{57}$$

If

$$\mathfrak{Y}^{(r)} = \begin{pmatrix} \mathfrak{Y}_1^{(r)} & \mathfrak{Y}_2^{(r)} \\ \mathfrak{Y}_3^{(r)} & \mathfrak{Y}_4^{(r)} \end{pmatrix},$$

then the elements of the matrices $\mathfrak{Y}_1^{(r)}$, $\mathfrak{Y}_2^{(r)}$, $\mathfrak{Y}_4^{(r)}$ are all bounded in absolute value by quantities depending only on K, m and $|\mathfrak{H}|$. Let

$$\mathfrak{Y}_3^{(r)} = \mathfrak{P}_r(\mathfrak{X}_1^{(r)} + \mathfrak{X}_2^{(r)}) \tag{58}$$

where $\mathfrak{X}_1^{(r)}$ is q_r-rowed symmetric and $\mathfrak{X}_2^{(r)}$ is skew symmetric. Also

$$\operatorname{abs} x_{kl}^{(r)} \leq c_8 (h_l^{(r)})^{-1}, \qquad (1 \leq k \leq l \leq q_r), \tag{59}$$

where $\mathfrak{X}_1^{(r)} = (x_{kl}^{(r)})$ and the elements of $\mathfrak{X}_2^{(r)}$ are bounded.

Each point in X_{at} is associated with a system of numbers $q_1, q_2, \cdots, q_f$, where $0 \leq q_r \leq [m/2]$ and $q_r \leq n_r$. Let $I(q_1, \cdots, q_f)$ be the integral extended over all points of X_{at} which have the same $q_1, \cdots, q_f$. Then

$$\int_X dv = \sum_{q_1=0}^{n_1} \sum_{q_2=0}^{n_2} \cdots \sum_{q_r=0}^{n_r} I(q_1, \cdots, q_f). \tag{60}$$

Now $I(0, 0, \cdots, 0)$ is bounded because then from (57), $|\mathfrak{T}^{(r)}|^{-m/2}$ is bounded for all r and from (58) all the elements of $\mathfrak{Y}^{(r)}$ are bounded so that $I(0, \cdots, 0)$ is an integral over a bounded domain with a bounded integrand. Let us now consider $I(q_1, \cdots, q_f)$, where not all $q_1, \cdots, q_f$ are zero. Let us split up the domain where $q_1, \cdots, q_f$ have a specified system of values not all zero into parallelpipeds

$$\begin{aligned} &\exp(-t g_k^{(r)}) < h_k^{(r)}/h_{k+1}^{(r)} \leq \exp(t(1 - g_k^{(r)})), \\ &\qquad\qquad (k = 1, \cdots, q_r - 1); \\ &\exp(-t g_{q_r}^{(r)}) < h_{q_r}^{(r)} \leq \exp(t(1 - g_{q_r}^{(r)})), \end{aligned} \tag{61}$$

where $c_1 = e^t > 1$ and $g_k^{(r)}$ for all $r = 1, \cdots, f$ and $k = 1, \cdots, q_r$, runs through all integers $0, 1, 2, \cdots$. What we have done is to split up the domain into a number of disjoint subsets. Let $f_k^{(r)} = g_k^{(r)} + \cdots + g_{q_r}^{(r)}$, $f_0^{(r)} = 0$, $k = 1, \cdots, q_r$. Then

$$c_6 \exp(-t f_k^{(r)}) \leq h_k^{(r)} \leq c_7 \exp(-t f_k^{(r)}). \tag{62}$$

Using (57) and (62) we get

$$|\mathfrak{T}^{(r)}|^{-m/2} \leq c_8 \exp[-\tfrac{1}{2} mt \sum_{k=0}^{q_r} f_k^{(r)}]. \tag{63}$$

The volume of every subdomain has the upper bound

$$c_9 \exp[\sum_{r=1}^{f} \sum_{k=0}^{q_r} t k f_k^{(r)}]. \tag{64}$$

Thus $I(q_1, \cdots, q_f)$ is majorised by

$$c_{10} \sum_{A} \exp(-At), \tag{65}$$

where $A = \sum_{r=1}^{f} \sum_{k=0}^{q_r} (\frac{1}{2}m - k) f_k^{(r)}$ and the summation in $g_1^{(r)}, \cdots$ runs through all non-negative rational integers. (65) is actually a convergent series if $m/2 - k > 0$ and diverges if $m = 2$ and $q_1 = q_2 = \cdots = q_f = 1$. This is the case of a binary decomposable form; for in this case $\mathfrak{S}$ is of the form (50), namely

$$\mathfrak{S}^{(r)} = \begin{pmatrix} 0 & p_r \\ p_r & g_r \end{pmatrix}. \tag{66}$$

In the statement of Theorem B we have excepted this case. Thus Theorem B is completely proved.

In the case of hermitian forms over K_0 the arguments are exactly similar. The difference arises in (57) where we have

$$N \mid \mathfrak{T}^{(r)} \mid^{-m/2} \leq c_7 \prod_{r=1}^{f} \prod_{k=1}^{q_r} (h_k^{(r)})^m \tag{67}$$

and so the analogue of the function A in (65) is

$$A = \sum_{r=1}^{f} \sum_{k=0}^{q_r} (m + 1 - 2k) f_k^{(r)}, \tag{68}$$

so that the fundamental domain has finite volume since $m + 1 > 2q_r$. We present an interesting proof of the finiteness in the case of decomposable hermitian forms; that is binary hermitian forms whose matrix is:

$$\mathfrak{S} = \begin{pmatrix} 0 & a \\ \bar{a} & b \end{pmatrix}, \qquad b = \bar{b}, \tag{69}$$

where a and b are integers of K_0. We show that in this case the measure is finite.

Let $\mathfrak{S}$ be a non-singular hermitian integral matrix in (69). If we make the transformation in K_0:

$$x_1 = -2(a - \bar{a})^{-1}(ax + \tfrac{1}{2}by), \qquad y_1 = y, \tag{70}$$

then $\mathfrak{S}$ goes into $\frac{1}{2}(\bar{a} - a)(x\bar{y} - y\bar{x})$ whose matrix is $\frac{1}{2}(\bar{a} - a)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. By (21) we can assume $(\bar{a} - a)/(2i)$ is totally positive, for otherwise we can interchange x and y or multiply by a suitable integer of K. It is easily seen that for proving the theorem it is enough to consider the units of $\frac{1}{2}(\bar{a} - a)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. This matrix we denote by $\mathfrak{S}_0$.

Let $\mathfrak{M} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ be a unit of $\mathfrak{S}_0$ in K_0. Then $\alpha\delta - \beta\gamma = \rho$, where ρ is a root of unity in K_0. The roots of unity in K_0 form a finite group [5] and so by considering those units for which $\alpha\delta - \beta\gamma = 1$ we shall only be considering a subgroup of finite index in the group of units. Let us call this subgroup the *principal group of units*. Let $\mathfrak{M} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$, where $\alpha, \beta, \gamma, \delta$ are integers such that $\alpha\delta - \beta\gamma = 1$, be in the principal group. Then

$$\alpha\bar{\gamma} = \gamma\bar{\alpha}, \qquad \beta\bar{\delta} = \delta\bar{\beta}, \qquad \alpha\bar{\delta} - \beta\bar{\gamma} = 1. \tag{71}$$

The most general solution of these equations is

$$\alpha/a = \beta/b = \gamma/c = \delta/d = \rho, \tag{72}$$

where ρ is a complex number of absolute value 1 and a, b, c, d are real. Since $\alpha, \beta, \gamma, \delta$ are integers of K_0 we see that a, b, c, d are integers of the totally real field K. Thus this principal group of units is none other than the restricted Hilbert modular group [2] in K.

The representation of this group is in a space of f complex variables $z_1, \cdots, z_f$ where

$$\mathfrak{S}^{(r)-1} \left\{ \frac{1}{z_r} \right\} = 2/(\bar{a} - a)(\bar{z}_r - z_r) > 0, \tag{73}$$

which is the space of the upper half planes of the f complex variables z_r. The group of units is represented in this space by

$$z_r \to \alpha^{(r)} z_r + \beta^{(r)}/\gamma^{(r)} z_r + \delta^{(r)}, \qquad (r = 1, \cdots, f), \tag{74}$$

and the volume element is

$$dv = \prod_{r=1}^{f} | \mathfrak{T}^{(r)} |^{-2} \, dz_1 \cdots dz_f, \tag{75}$$

where $\mathfrak{T}^{(r)} = 2(\bar{a} - a)^{-1}(\bar{z}_r - z_r)$ and $dz_r = 2dx_r dy_r$, $z_r = x_r + iy_r$. The measure of volume is $dx_1 \cdots dx_f dy_1 \cdots dy_f / {y_1}^2 \cdots {y_f}^2$. From the work of Blumenthal [2] it is known that (74) has a fundamental domain bounded by a finite number of surfaces and by Siegel's work [11], this fundamental domain has a finite volume.

7. Decomposable binary forms. In this section we shall show that in the case of a binary decomposable quadratic form a fundamental domain has actually infinite volume. From the considerations of Section **4** it is enough if we consider the form $\mathfrak{x}'\mathfrak{S}\mathfrak{x}$ where

$$\mathfrak{S} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{76}$$

The system $(\mathfrak{S})$ is $(\mathfrak{S}^{(1)}, \cdots, \mathfrak{S}^{(f)})$, where $\mathfrak{S}^{(r)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

Let $\mathfrak{U} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ be a unit of $\mathfrak{S}$. It is then easy to see that $\mathfrak{U}$ has one of the two forms

$$\begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}, \quad \begin{pmatrix} 0 & \epsilon \\ \epsilon^{-1} & 0 \end{pmatrix}, \tag{77}$$

where ϵ is an algebraic unit in K. Let Γ_0 be the group of units of $\mathfrak{S}$ and Γ_1 the subgroup of units of the form $\begin{pmatrix} \epsilon & 0 \\ 0 & \epsilon^{-1} \end{pmatrix}$. Then Γ_0/Γ_1 is of order 2.

Now the space J can be easily seen to be the space of f real variables $u_1, \cdots, u_f$ for which

$$u_i > 0, \qquad (i = 1, \cdots, f). \tag{78}$$

The representation of Γ_1 in this space is

$$u_i \to \epsilon^{(i)2} u_i, \qquad (i = 1, \cdots, f). \tag{79}$$

The volumes of the fundamental domains for Γ_0 and Γ_1 are in the ratio 1 : 2, so that it is enough if we consider Γ_1.

From Dirichlet's theorem we see that any unit ϵ may be written as $\epsilon = \pm \epsilon_1^{a_1} \cdots \epsilon_{f-1}^{a_{f-1}}$. Let us make the transformation

$$\log u_i = v_i, \qquad (i = 1, \cdots, f) \tag{80}$$

(since $u_i > 0$ this is permissible). We have then the whole of the space V of $v_1, \cdots, v_f$. The transformations (79) become

$$v_i \to v_i + a_1 \lambda_1^{(i)} + \cdots + a_{f-1} \lambda_{f-1}^{(i)}, \tag{81}$$

where $\lambda_j^{(i)} = \log(\text{abs } \epsilon_j^{(i)})^2$, $i = 1, \cdots, f$; $j = 1, \cdots, f-1$. Also since $N\epsilon^2 = 1$ we have for any unit ϵ_j

$$\sum_{i=1}^{f} \lambda_j^{(i)} = 0, \tag{82}$$

which means that the mappings (81) preserve the planes

$$v_1 + \cdots + v_f = \text{a constant}. \tag{83}$$

It is known [5] that in an algebraic number field K the matrix

$$(\lambda_j^{(i)}), \qquad i = 1, \cdots, f; \; j = 1, \cdots, f-1, \tag{84}$$

is of rank $f-1$. (This is none other than the assertion that the regulator

is non-vanishing.) This means that every point $v_1, \cdots, v_f$ in V may be written

$$v_i = \eta_0/f + \eta_1\lambda_1^{(i)} + \cdots + \eta_{f-1}\lambda_{f-1}^{(i)}, \tag{85}$$

where $\eta_0 = v_1 + \cdots + v_f$ and $\eta_1, \cdots, \eta_{f-1}$ are real numbers. Consider now the plane $v_1 + \cdots + v_f = 0$. Because of (81) the points $(\eta_i\lambda_i^{(1)}, \cdots, \eta_i\lambda_i^{(f)})$, $i = 1, \cdots, f-1$, lie in this plane. Consider rectangular coordinates in this plane with origin as centre. Consider the $f-1$ points

$$(\lambda_i^{(1)}, \cdots, \lambda_i^{(f)}), \qquad i = 1, \cdots, f-1. \tag{86}$$

Because of (82) these points together with the origin form the vertices of a $(f-1)$ simplex. This simplex can be completed into an $(f-1)$ dimensional parallelepiped. We make this a fundamental parallelepiped by prescribing that the boundary points lying on the faces of this parallelepiped containing the origin shall belong to the fundamental region. We call this F_1. Choose now a fixed direction through the origin and not lying in this plane. Through every point of F_1 draw lines parallel to this direction. The resulting point set F can easily be verified by the method of Blumenthal [2] to be a fundamental domain for (79).

The volume element in the $u_1, \cdots, u_f$ space is $du_1 \cdots du_f/u_1 \cdots u_f$. That in the $v_1, \cdots, v_f$ space is the Euclidean volume element $dv_1 \cdots dv_f$. The Euclidean volume of F is obviously infinite. Thus the

THEOREM C. *The volume of a fundamental domain for the units of a binary decomposable quadratic form over a totally real field is infinite.*

In the case of rational decomposable binary forms this theorem is at once obvious from Theorem D below, because the number of units of a binary decomposable form over the field of rationals is finite. In the case of algebraic number fields this result does not follow from Theorem D.

8. Volume of the space *J*. In this section we shall prove

THEOREM D. *The volume of the space J measured with the volume element* (22) *is infinite.*

This is also true of the space (31) in case of hermitian forms. We shall prove it in the case $f = 1$ only. The general result follows at once by taking topological products.

It is evident that we can assume that $\mathfrak{S} = \begin{pmatrix} \mathcal{E}_n & \mathfrak{O} \\ \mathfrak{O} & -\mathcal{E}_{m-n} \end{pmatrix}$. Let us put $n = p$ and $m - n = q$ so that $p + q = m$. The J space is now

$$\mathcal{E} - \mathfrak{Y}'\mathfrak{Y} > 0, \tag{87}$$

where $\mathcal{E}$ is the unit matrix of order p and $\mathfrak{Y}$ is a real matrix of p columns and q rows. The invariant volume element is

$$dv = |\mathcal{E} - \mathfrak{Y}'\mathfrak{Y}|^{-\frac{1}{2}(p+q)} d\mathfrak{Y}. \tag{88}$$

Before proving Theorem D we make the following remarks.

If $\mathfrak{S}$ is a positive definite symmetric matrix and $\mathfrak{S} = \begin{pmatrix} \mathfrak{S}_1 & \mathfrak{S}_2 \\ \mathfrak{S}'_2 & \mathfrak{S}_3 \end{pmatrix}$, where $\mathfrak{S}_1$, $\mathfrak{S}_3$ are symmetric, then

$$|\mathfrak{S}| \leq |\mathfrak{S}_1|\,|\mathfrak{S}_3|. \tag{89}$$

In the second place if $\mathcal{E} - \mathfrak{Y}'\mathfrak{Y} > 0$ then $|\mathcal{E} - \mathfrak{Y}'\mathfrak{Y}| \leq 1$. This is easy, for if $\mathfrak{Y} = (y_{kl})$, then the diagonal elements of $\mathcal{E} - \mathfrak{Y}'\mathfrak{Y}$ are

$$1 - \sum_{t=1}^{q} y_{tk}^2 > 0 \qquad (k = 1, \cdots, p).$$

Also these diagonal elements are < 1. From (89) follows the result.

We prove Theorem D by induction on p. When $p = 1$, the space is

$$J_1: \quad 1 - y_1^2 - \cdots - y_q^2 > 0 \tag{90}$$

which is the interior of the unit sphere in the q dimensional space. It is then well known that

$$\int_{J_1} (1 - y_1^2 - \cdots - y_q^2)^{-\frac{1}{2}(q+1)} dy_1 \cdots dy_q \tag{91}$$

diverges for $q > 0$. Theorem D is thus true for $p = 1$. Let us assume it proved for $p - 1$ instead of p. Let $\mathfrak{Y}_1$ be the first $p - 1$ columns of $\mathfrak{Y}$ and $\mathfrak{x}$ the last column so that $\mathfrak{Y} = (\mathfrak{Y}_1 \mathfrak{x})$. Then

$$\mathcal{E} - \mathfrak{Y}'\mathfrak{Y} = \begin{pmatrix} \mathcal{E} - \mathfrak{Y}_1'\mathfrak{Y}_1 & -\mathfrak{Y}_1'\mathfrak{x} \\ -\mathfrak{x}'\mathfrak{Y}_1 & 1 - \mathfrak{x}'\mathfrak{x} \end{pmatrix} > 0, \tag{92}$$

which shows that $\mathcal{E} - \mathfrak{Y}_1'\mathfrak{Y}_1 > 0$, $1 - \mathfrak{x}'\mathfrak{x} > 0$. Using (89) we get

$$|\mathcal{E} - \mathfrak{Y}'\mathfrak{Y}| \leq |\mathcal{E} - \mathfrak{Y}_1'\mathfrak{Y}_1|\,(1 - \mathfrak{x}'\mathfrak{x}) \tag{93}$$

and hence

$$\int_{J_p} |\mathcal{E} - \mathfrak{Y}'\mathfrak{Y}|^{-\frac{1}{2}(p+q)} d\mathfrak{Y} \geq \int_{J_p} |\mathcal{E} - \mathfrak{Y}_1'\mathfrak{Y}_1|^{-\frac{1}{2}(p+q-1)} (1 - \mathfrak{x}'\mathfrak{x})^{-(p+1)/2} d\mathfrak{Y}_1 d\mathfrak{x}. \tag{94}$$

If $\mathfrak{Y}_1$ is a real matrix of $p-1$ columns and q rows, then it is possible to choose a column vector $\mathfrak{x}$ such that

$$\mathcal{E}-\mathfrak{Y}'\mathfrak{Y}>0, \qquad \mathcal{E}-\mathfrak{Y}_1'\mathfrak{Y}_1>0, \qquad \mathfrak{Y}=(\mathfrak{Y}_1\mathfrak{x}). \tag{95}$$

This is seen from the identity

$$(\mathcal{E}-\mathfrak{Y}'\mathfrak{Y})=\begin{pmatrix}\mathcal{E}-\mathfrak{Y}_1'\mathfrak{Y}_1 & \mathfrak{O}\\ \mathfrak{O} & \mathfrak{Q}\end{pmatrix}\begin{bmatrix}\mathcal{E} & -(\mathcal{E}-\mathfrak{Y}_1'\mathfrak{Y}_1)^{-1}\mathfrak{Y}_1'\mathfrak{x}\\ \mathfrak{O} & 1\end{bmatrix}, \tag{96}$$

where

$$\mathfrak{Q}=1-\mathfrak{x}'\mathfrak{x}-\mathfrak{x}'\mathfrak{Y}_1(\mathcal{E}-\mathfrak{Y}_1'\mathfrak{Y}_1)^{-1}\mathfrak{Y}_1'\mathfrak{x}>0. \tag{97}$$

Now (97) is necessary and sufficient for $\mathcal{E}-\mathfrak{Y}'\mathfrak{Y}>0$. Also (96) implies that $1-\mathfrak{x}'\mathfrak{x}>0$, so that if J_{p-1} is the space

$$\mathcal{E}-\mathfrak{Y}_1'\mathfrak{Y}_1>0, \tag{98}$$

then from (94) we get

$$\int_{J_p}|\mathcal{E}-\mathfrak{Y}'\mathfrak{Y}|^{-(p+q)/2}d\mathfrak{Y}\geq\int_{J_{p-1}}|\mathcal{E}-\mathfrak{Y}_1'\mathfrak{Y}_1|^{-\frac{1}{2}(p+q-1)}d\mathfrak{Y}_1, \tag{99}$$

which now completes the induction.

It is possible to obtain Theorem D from topological considerations alone. Siegel [10] has shown that the space H of positive solutions of

$$\mathfrak{H}\mathfrak{S}^{-1}\mathfrak{H}=\mathfrak{S} \tag{100}$$

can be considered as the homogeneous space of the cosets of a Lie group $\mathfrak{G}$ by a closed maximal compact subgroup $\mathfrak{C}$. From the existence of invariant measures in $\mathfrak{G}$ and $\mathfrak{G}/\mathfrak{C}$ it is evident that $\mathfrak{G}/\mathfrak{C}$ has infinite measure for otherwise, $\mathfrak{C}$ being compact, $\mathfrak{G}$ has finite measure and therefore is compact, which is absurd. Theorem D is also implicit in Theorem 1 of Siegel [10]. For we can show that H contains a geodesic sphere of arbitrarily large radius and since the volume of a geodesic sphere of radius r is $>v(r)$ where $v(r)$ is the volume of a Euclidean sphere of pq dimensions and of radius r, the theorem D becomes obvious.

INSTITUTE FOR ADVANCED STUDY,
PRINCETON, NEW JERSEY.

BIBLIOGRAPHY.

[1] E. Artin, *Galois Theory*, Notre Dame, 1944.

[2] O. Blumenthal, "Über Modulfunktionen von mehreren Veränderlichen," *Mathematische Annalen*, Bd. 56 (1903), pp. 509-548.

[3] Dirichlet-Dedekind, *Vorlesungen über Zahlentheorie*, Vierte Auflage, p. 149.

[4] R. Fricke, "Zur gruppentheoretische Grundlegung der automorphen Funktionen," *Mathematische Annalen*, vol. 42 (1893), pp. 564-594.

[5] E. Hecke, *Theorie der algebraischen Zahlen*, Leipzig, 1923.

[6] P. Humbert, "Théorie de la réduction des formes quadratiques définies positives dans un corps algébrique K fini," *Commentarii Mathematici Helvetici*, vol. 12 (1940), pp. 263-306.

[7] ———, "Réduction des formes quadratiques dans un corps algébrique fini," *ibid.*, vol. 23 (1949), pp. 50-63.

[8] C. L. Siegel, "Einheiten quadratischer Formen," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 13 (1940), pp. 209-239.

[9] ———, "On the theory of indefinite quadratic forms," *Annals of Mathematics*, vol. 45 (1944), pp. 577-622.

[10] ———, "Some remarks on discontinuous groups," *ibid.*, vol. 46 (1945), pp. 708-718.

[11] ———, "The volume of the fundamental domain for some infinite groups," *Transactions of the American Mathematical Society*, vol. 39 (1936), pp. 209-218.

THE EXISTENCE AND LIMIT BEHAVIOR OF THE ONE-DIMENSIONAL SHOCK LAYER.*

By DAVID GILBARG.

Introduction. We consider steady, one-dimensional flows of a viscous, heat-conducting fluid which approach finite limit values at $x = +\infty$ and $x = -\infty$. Such flows display the character of a shock wave (for small viscosity, μ, and heat conductivity, λ), in that they differ sensibly from their end states at $x = \pm\infty$ only in a small interval of rapid transition. In analogy with the classical boundary layer, and also to distinguish these flows from the shock waves which belong properly to the theory of ideal fluids, we follow Weyl [1] in naming such a flow a *shock layer*. The one-dimensional shock layer is in certain respects the prototype of all shock phenomena and has therefore been studied widely, with particular emphasis on the problem of thickness of the shock front [2, 3, 5, 6, 8]. However, basic problems concerning these flows, such as those of existence, and limit behavior for small λ, μ, remain open. Their solution, which we consider here, is a step towards placing on a sound basis the relation between the theories of real and ideal fluids.

The general problem of existence of the shock layer for a fluid with given λ, μ, and with the preassigned end states, has been studied inconclusively by Rayleigh [4] and Weyl [1]. Until now, the existence of the shock layer seems to have been definitely proved only for an exceptional set of ideal gases for which a postulated relation between λ, μ, and the specific heat at constant pressure,[1] permits explicit integration of the equations of motion; (Becker [2], also [5, 6]). We succeed here in obtaining an essentially complete solution of the existence problem by proving the existence and uniqueness of the shock layer for the general class of fluids considered by Weyl, with λ, μ arbitrary functions of the state, and for arbitrary end states satisfying the shock relations (Theorem 1). This result, therefore, establishes for general fluids an exact correspondence between the steady one-dimensional shock waves and the shock layers.

* Received June 3, 1950; revised November 13, 1950.
Prepared under Navy Contract N6onr-180, Task Order V, with Indiana University.

[1] Namely, Prandtl number $c_p\mu/\lambda = 3/4$.

The problem of the limit behavior of the shock layers for small λ and μ is that of providing a rigorous proof for the statement, generally accepted on physical grounds, that any shock wave in an ideal fluid is the limit of the corresponding shock layers, and, conversely, that the shock layers (with the same end states) approach a shock wave in the limit as $\lambda, \mu \to 0$.[2] Mathematically, this problem belongs to the class of perturbation problems for differential equations in which the vanishing of certain parameters, (λ, μ in this case), reduces the order of the equations. A complete solution to the problem is given in Theorems 2 and 5 for the same general fluids as in the existence theorem. The limit behavior makes apparent the unequal dependence of the flows on viscosity and heat conductivity in the fact that as $\lambda \to 0$ (with μ fixed), the shock layers converge to a continuous thermally non-conducting shock layer (Theorem 3), whereas for $\mu \to 0$ (with λ fixed), they converge to a (generally) *discontinuous* non-viscous shock layer (Theorem 4).

We remark that the methods used here provide easily estimated bounds on the thickness of the shock front, although not the precise determinations sought by previous writers [2, 3, 5, 6] on this subject.

1. Existence and uniqueness of the shock layer. The equations of motion for a steady one-dimensional flow of a viscous heat conducting fluid are:[3]

$$\rho u = \text{constant} = c_1; \tag{1}$$

$$\rho u^2 + p - \mu u_x = \text{constant} = c_2 \tag{2}$$

$$\rho u(\tfrac{1}{2}u^2 + e + p/\rho) - \mu u u_x - \lambda \theta_x = \text{const.} = c_3. \tag{3}$$

These equations express respectively the conservation conditions for mass, momentum, and energy. The coefficient of viscosity, μ (which combines the two viscosity coefficients appearing in the general Navier-Stokes equations), and the heat conductivity, λ, are in general functions of the thermodynamic state, and will here be considered in this generality. The other quantities appearing in the equations are the density ρ, pressure p, velocity u, internal energy e per unit mass, and temperature θ. We assume that the variables θ and $\tau = 1/\rho =$ specific volume, fix uniquely the thermodynamic state of the fluid, and that λ, μ, p, and e are sufficiently smooth (e. g. twice differentiable), functions of these variables.

---

[2] This limit behavior is apparent in the explicit solution obtained by Becker [2]; however, in this case, λ and μ are connected by the relation, $c_p\mu/\lambda = 3/4$, and do not approach zero independently. For discussion of the problem of limit behavior, see, for example, [7] pp. 135-138, also [8], pp. 218-222.

[3] See, for example, [7] p. 134; a complete derivation is contained in [9].

Eliminating u from equations (1)-(3), we obtain,

$$\lambda(d\theta/dx) = b[e(\tau,\theta) - \tfrac{1}{2}b^2(\tau-a)^2 - c] \equiv L(\tau,\theta)$$

(E)

$$\mu(d\tau/dx) = 1/b[p(\tau,\theta) + b^2(\tau-a)] \equiv M(\tau,\theta),$$

where,

$$a = c_2/c_1^2, \qquad b = c_1, \qquad c = c_3/c_1 - c_2^2/2c_1^2,$$

and $e = e(\tau,\theta)$, $p = p(\tau,\theta)$ are (given) equations of state of the fluid. With certain minor modifications, this is the form of the equations of motion used by both Weyl [1] and Becker [2].

We enumerate now a set of conditions on the functions $L(\tau,\theta)$, $M(\tau,\theta)$ which will be sufficient to prove the desired existence and limit theorems. In Section **3**, it will be verified that the corresponding functions for the general fluids of Weyl, and, in particular, for polytropic gases, satisfy these conditions. We assume:

The domain of definition is a set of points $Z = (\tau,\theta)$, hereafter called the Z-plane, forming a simply connected region in the quadrant $\tau > 0$, $\theta > 0$. In this region, we have:

(A) $L_\theta > 0$, $M_\theta > 0$.

(B) There are two curves, L and M, on which $L(\tau,\theta) = 0$ and $M(\tau,\theta) = 0$, respectively, and which intersect in two points, $Z_0 = (\tau_0,\theta_0)$, $Z_1 = (\tau_1,\theta_1)$, $(\tau_0 > \tau_1)$; (τ_0,θ_0) and (τ_1,θ_1) are the only simultaneous solutions of $L(\tau,\theta) = 0$ and $M(\tau,\theta) = 0$.

(C) $L_\tau > 0$ on L for $\tau_1 \leq \tau \leq \tau_0$.

(D) $L_\tau/L_\theta < M_\tau/M_\theta$ at Z_0; $L_\tau/L_\theta > M_\tau/M_\theta$ at Z_1.

It is an immediate consequence of (A) that L and M are representable as single valued functions of τ, namely, $\theta = l(\tau)$, $\theta = m(\tau)$, respectively. By virtue of (C), $l(\tau)$ is monotonically decreasing in the interval $\tau_1 \leq \tau \leq \tau_0$, and thus $\theta_1 > \theta_0$. From (B) and (D) it follows $m(\tau) > l(\tau)$ for $\tau_1 < \tau < \tau_0$.

The closed curve formed by the arcs $L^* \subset L$ and $M^* \subset M$ which join Z_0 and Z_1 bounds a simply connected region R of the Z-plane (Fig. 1). We conclude from (A) that everywhere in R, $L(\tau,\theta) > 0$ and $M(\tau,\theta) < 0$.

The preceding geometric facts contain all the information that is needed about the (τ,θ) phase plane for the proofs which follow. More general conditions could be obtained, but would be of no interest for the ensuing applications.

Consider the points $Z_0(\tau_0,\theta_0)$ and $Z_1(\tau_1,\theta_1)$. These satisfy the shock conditions,

$$p_0 + b^2\tau_0 = p_1 + b^2\tau_1 = b^2 a \tag{4}$$

$$e_0 - b^2/2(\tau_0 - a)^2 = e_1 - b^2/2(\tau_1 - a)^2 = c. \tag{5}$$

Z_0 and Z_1 then represent possible initial and final states, respectively, of a normal shock wave of an ideal fluid having the same equations of state as the given fluid. A solution, $S(x) = (\tau(x), \theta(x))$, $(-\infty < x < +\infty)$, of equations (E), (for given $\lambda(\tau, \theta)$, $\mu(\tau, \theta)$), will be called a *shock layer* if

$$\lim_{x \to -\infty} S(x) = Z_0, \qquad \lim_{x \to \infty} S(x) = Z_1.$$

The corresponding *shock layer curve* in the Z-plane is the integral curve

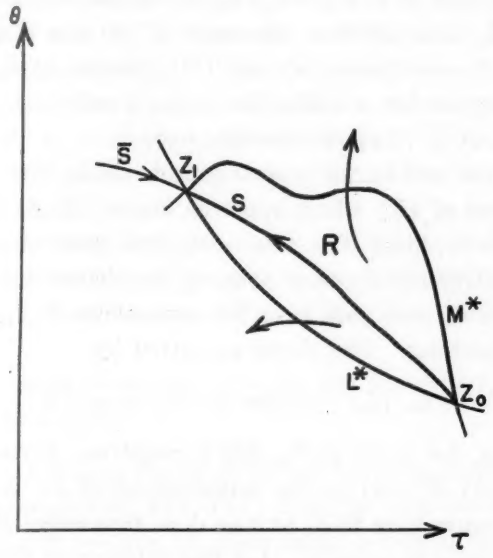

FIG. 1.

represented by the set of "equivalent" shock layers, $S(x + h)$, (h = constant). A shock layer will be called *parametrized* if a particular representative of this class is designated, and, unless so noted, equivalent shock layers will be considered identical.

We observe that if an integral of (E) has the limits Z'_0, Z'_1 at $x = -\infty$, $+\infty$, respectively, then Z'_0 Z'_1 must satisfy the shock conditions (4), (5). It is necessary, therefore, to restrict the definition of shock layer, as we have done, to end states defining a normal shock wave.

We consider first the question of existence and uniqueness of the shock layer for arbitrary positive $\lambda = \lambda(\tau, \theta)$, $\mu = \mu(\tau, \theta)$.

By condition (B), Z_0 and Z_1 are the exclusive singular points of the system (E).[4] We prove that Z_0 is an unstable node and Z_1 a saddle point of (E).[5] For, the characteristic equation of the system (E) at Z_0 and Z_1 is

$$(6) \qquad 0 = \begin{vmatrix} L_\theta/\lambda - \kappa & L_\tau/\lambda \\ M_\theta/\mu & M_\tau/\mu - \kappa \end{vmatrix}$$
$$= \kappa^2 - (M_\tau/\mu + L_\theta/\lambda)\kappa + (M_\theta L_\theta/\lambda\mu)(M_\tau/M_\theta - L_\tau/L_\theta),$$

where the values λ, μ, L_τ, M_θ, etc., are to be taken at Z_0 and Z_1. The discriminant of this equation is

$$[(M_\tau/\mu - L_\theta/\lambda)^2 + 4(M_\theta L_\tau/\lambda\mu)]_{z=z_0,z_1};$$

by conditions (A) and (C), M_θ, $L_\tau > 0$, so that the discriminant is positive at both Z_0 and Z_1, and therefore the roots of (6) are real. The constant terms of (6), is, by conditions (A) and (D), positive at Z_0, negative at Z_1. The latter point is therefore a saddle, the former a node, and, since M_τ, $L_\theta > 0$ (conditions (A), (C), (D)), an unstable node.

It follows from well-known general considerations that there are exactly two integral curves of (E) which approach the saddle Z_1 as $x \to +\infty$, and exactly two which approach it as $x \to -\infty$, these pairs corresponding to the negative and positive roots, respectively, of the characteristic equation (6). The two members of each pair have the same slope at Z_1, but approach it from opposite directions. The slopes are given by

$$-L_\tau/(L_\theta - \kappa\lambda) = -(M_\tau - \kappa\mu)/M_\theta,$$

and, in particular, for $\kappa < 0$ at Z_1, this is negative. Hence, by considering the sign of $L(\tau, \theta)/M(\tau, \theta)$ in the neighborhood of Z_1, it is seen that one of the solutions converging to Z_1 as $x \to +\infty$ approaches it from the region R, in which the ratio is negative. Let this solution be designated by $S(x)$. We establish that $S(x)$ is a shock layer. For, consider the integral curves of (E) which pass through the points of M^* and L^*. On M^*, $M(\tau, \theta) = 0$, $L(\tau, \theta) > 0$, so that all integral curves have vertical tangent vectors, and are directed outwards from R for increasing x. Similarly, on L^*, all solutions have zero slope, with $L(\tau, \theta) = 0$, $M(\tau, \theta) < 0$, and are directed outwards from R, since the slope of L^* is negative (by conditions (A) and (C)). Thus, for *decreasing* x, all integral curves of (E) passing through L^* and M^*

[4] This makes clear that the shock transition from Z_0 to Z_1 cannot (strictly) occur in a finite x-interval, since the singular points of (E) can be approached only for $x = \pm\infty$.

[5] This is proved by Weyl in the (τ, p) plane ([1], Theorem 6).

are directed *into* R. In R itself, all integral curves have negative slope and are traversed for decreasing x in the direction of increasing τ and decreasing θ. Let us now follow $S(x)$ for decreasing x. By virtue of the preceding, $S(x)$ cannot intersect L^* or M^* between Z_0 and Z_1. Since there are no other singular points of (E), and since $S(x)$ is monotonic while in R, and cannot terminate inside R, it must approach the node Z_0 as $x \to -\infty$. This proves the existence of a shock layer $S(x) = (\tau(x), \theta(x))$.

To prove the uniqueness of the corresponding shock layer curve S, we show that no other integral curve of (E) can join Z_0 and Z_1. For Z_0, being an unstable node, which can be approached by an integral curve only as $x \to -\infty$, the only solution other than S which could join Z_0 and Z_1 would have to be the second integral curve, call it $\bar{S}$, which enters Z_1 as $x \to +\infty$. This is seen to be impossible as follows. If $\bar{S}$ is also a shock layer curve, then the arcs S and $\bar{S}$ form a closed curve bounding a simply connected region G in the Z-plane. One of the two integral curves of (E) which approach Z_1 as $x \to -\infty$ enters this region. It cannot terminate in G, since Z_0 and Z_1 are the exclusive singular points of (E); nor can it approach asymptotically a limit cycle in G, for this in turn would have to contain in its interior a singular point of (E).[6] The only possibility remaining[6] is that the curve intersect S or $\bar{S}$, thereby contradicting the uniqueness of the integral curves of (E). This contradiction completes the uniqueness proof.

Thus, under assumptions (A)-(D), we can state the following existence and uniqueness theorem.

THEOREM 1. *Let $Z_0 = (\tau_0, \theta_0)$ and $Z_1 = (\tau_1, \theta_1)$ be states of a fluid which satisfy the shock conditions,* (4), (5); *then, for any viscosity $\mu = \mu(\tau, \theta)$, and heat conductivity $\lambda = \lambda(\tau, \theta)$, there exists a unique shock layer joining Z_0 to Z_1.*

2. Limit behavior of the shock layer as $\lambda \to 0$, $\mu \to 0$. For given $\lambda(Z)$, $\mu(Z)$, let the shock layers of the preceding section be designated by $S(x; \lambda, \mu) = (\tau(x; \lambda, \mu), \theta(x; \lambda, \mu))$, and the associated layer curve by $S(\lambda, \mu)$. We assume now concerning $\lambda(Z)$ and $\mu(Z)$ that these coefficients depend on parameters (independent of τ, θ) in such a way that λ and μ can independently be made arbitrarily small in R. In the following, when bounds are placed on λ, μ, or λ/μ, these bounds are to be understood as holding throughout the region R.

---

[6] Kamke, *Differentialgleichungen Reeller Funktionen*, Chelsea, New York (reprinted 1947), pp. 216, 220-222.

It is our chief object in this section to establish that, as λ and μ approach zero (in whatever manner), the corresponding shock layers approximate a shock wave. In a sense, this "justifies" the usual shock wave theory by showing it to be the limit case of the corresponding theory of real fluids. In the details, it is shown that the limits as $\lambda, \mu \to 0$ exist in the sense of function limits. This requires the individual shock layers to be so parametrized—since they are fixed only to within an x-translation—that the appropriate limits exist under this parametrization. We consider first the double limit, under the same assumptions as in Theorem 1.

THEOREM 2. *Let $Z_0 = (\tau_0, \theta_0)$ and $Z_1 = (\tau_1, \theta_1)$ be initial and final states of the shock wave, $Z = Z_0$, $-\infty < x < \xi$, $Z = Z_1$, $\xi < x < \infty$. Then, the corresponding family of shock layers, $S(x; \lambda, \mu)$, if suitably parametrized, approaches this shock wave as $\lambda \to 0$ and $\mu \to 0$ independently; that is,*

$$\lim_{\substack{\lambda \to 0 \\ \mu \to 0}} S(x; \lambda, \mu) = Z_0, \quad -\infty < x < \xi,$$
$$= Z_1, \quad \xi < x < \infty,$$

the convergence being uniform in every closed interval not containing $x = \xi$.

Proof.[7] For $\epsilon > 0$, let $R(\epsilon)$ be the subregion of R outside the circles of radius ϵ about Z_0, Z_1. Consider the equations (E) for any integral curve in R in the form,

$$\text{(E)} \qquad d\theta/dx = L(\tau, \theta)/\lambda, \qquad d\tau/dx = M(\tau, \theta)/\mu,$$

and take the difference,

$$d(\theta - \tau)/dx = L(\tau, \theta)/\lambda + |M(\tau, \theta)|/\mu > 0$$

Define $\eta = \text{Max}(\lambda, \mu)$ in R, and let $C(\epsilon)$ be a constant for which

$$L(\tau, \theta) + |M(\tau, \theta)| \geq C(\epsilon) > 0, \qquad (\tau, \theta) \in R(\epsilon);$$

such a $C(\epsilon)$ exists, since $L(\tau, \theta) > 0$ on M^*, and $|M(\tau, \theta)| > 0$ on L^* (except at Z_0, Z_1), and are also positive in $R(\epsilon)$. Then, we have for any integral curve in $R(\epsilon)$,

$$(7) \qquad d(\theta - \tau)/dx > C(\epsilon)/\eta > 0$$

Now consider any shock layer $S(x; \lambda, \mu)$, and let (τ_M, θ_M) and (τ_m, θ_m) designate the value of $S(x; \lambda, \mu)$ where it intersects the ϵ circles about Z_1

[7] The author is indebted to E. Hopf for this proof, which strengthens and simplifies the author's original proof.

and Z_0 respectively, with $S(x_0;\lambda,\mu) = (\tau_m, \theta_m)$ and $S(x_1;\lambda,\mu) = (\tau_M, \theta_M)$. We have, for $x_0 \leq x \leq x_1$,

$$\tau_0 > \tau_m \geq \tau(x;\lambda,\mu) \geq \tau_M > \tau_1;$$
$$\theta_0 < \theta_m \leq \theta(x;\lambda,\mu) \leq \theta_M < \theta_1.$$

From inequality (7) for $S(x;\lambda,\mu)$, it follows,

$$(8) \quad x_1 - x_0 \leq \eta/C(\epsilon)[\theta_M - \tau_M - (\theta_m - \tau_m)] < \eta/C(\epsilon)[\theta_1 - \theta_0 + \tau_0 - \tau_1],$$

Thus, given ϵ and δ, and $S(x;\lambda,\mu)$ for which

$$\eta = \mathrm{Max}(\lambda,\mu) \leq \delta C(\epsilon)/(\theta_1 - \theta_0 + \tau_0 - \tau_1),$$

the simultaneous inequalities,

$$(9) \qquad |\tau(x;\lambda,\mu) - \tau_{0,1}| > \epsilon, \qquad |\theta(x;\lambda,\mu) - \theta_{0,1}| > \epsilon,$$

can hold only when the values of $S(x;\lambda,\mu)$ lie in $R(\epsilon)$, and therefore in an interval of length less that $x_1 - x_0 < \delta$. In other words, by making heat conductivity and viscosity sufficiently small, the width of the shock front can be made arbitrarily small. To complete the proof of the theorem, one need only observe that the solutions $S(x;\lambda,\mu)$ can be so parametrized that the intervals (x_0, x_1) contain $x = \xi$. The convergence of the $S(x;\lambda,\mu)$ to the discontinuous shock is then uniform outside of every open interval containing $x = \xi$.

We note that with arbitrary parametrization, either the double limit is a shock wave, as above, or does not exist in an entire interval, or is a constant, Z_0 or Z_1. Hence, among all parametrizations, the only non-trivial limits of shock layers, as $\lambda, \mu \to 0$ independently, are the shock waves.

We note also that (8) gives a (probably very poor) upper bound on the width of the shock transition in the sense of inequality (9). Similarly, lower bounds as well can be obtained from reverse inequalities of the type (7). It is possible that, by taking other linear combinations of the two equations (E), and thereby obtaining inequalities analogous to (7), more satisfactory bounds can be achieved.

We turn now to the single limit problems, $\lim_{\lambda\to 0} S(x;\lambda,\mu)$ and $\lim_{\mu\to 0} S(x;\lambda,\mu)$, for fixed $\mu(Z)$ and $\lambda(Z)$, respectively. First we prove,

LEMMA 1. *Let G be any open neighborhood of the closed arc L^*. Then for sufficiently small values of λ/μ, all shock layer curves $S(\lambda,\mu)$ lie entirely in G.*

Proof. Let $\bar{L}$ be an arc of bounded negative slope with endpoints $\neq Z_0$, Z_1 on M^*, and lying so close to L^* that the subregion of R contained between L^* and $\bar{L}$ contains only points of G. Such a curve $\bar{L}$ can always be found by virtue of the monotonicity of L^* between Z_0 and Z_1. Let D be the subregion of R bounded by $\bar{L}$ and M^*; D contains the complement of G in R. On $\bar{L}$, $L(\tau, \theta) \geq k_1 > 0$, $|M(\tau, \theta)| \leq k_2$, and $|\text{slope } \bar{L}| \leq N$, where k_1, k_2, and N are suitable positive constants. Let now $\mu/\lambda > (k_2/k_1)N$; then since the slopes of the corresponding integral curves of (E) satisfy on $\bar{L}$ the inequality,

$$-d\theta/d\tau = -(\mu/\lambda)(L(\tau, \theta)/M(\tau, \theta)) \geq (\mu/\lambda)(k_1/k_2) > N,$$

we have that these integral curves must be directed into D. Thus, any integral curve of (E) for which $\mu/\lambda > (k_2/k_1)N$ and which contains a point of D cannot intersect $\bar{L}$ beyond this point (i. e. for increasing x), and hence cannot pass through Z_1. As a consequence, if $S(\lambda, \mu)$ is a shock layer curve for which $\lambda/\mu < k_1/k_2N$, $S(\lambda, \mu)$ must lie entirely in the region G (indeed, in the smaller region between L^* and $\bar{L}$).

Similar considerations apply in case of small μ/λ. An interesting difference arises from the possibility that M^* is not monotonic. If the arc is monotonic, then exactly the same argument as above applies, with but evident verbal changes. If, however, M^* is not monotonic between Z_0 and Z_1, consider then the arc $\bar{M}^*$ defined by the monotonic function in $[\tau_1, \tau_0]$,

$$\theta = \bar{m}(\tau) = \operatorname*{Min}_{\tau_1 \leq t \leq \tau} m(t); \qquad \tau_1 \leq \tau \leq \tau_0.$$

This arc (Fig. 2) joining Z_1 and Z_0[8] encloses between it and L^* a subregion of R in which, we assert, all shock layer curves must be contained. For, if $\theta = \sigma(\tau)$ is the equation of any shock layer, we have, for every τ in $[\tau_1, \tau_0]$, and some t such that $\tau_1 \leq t \leq \tau$,

$$\sigma(\tau) \leq \sigma(t) \leq m(t) = \bar{m}(\tau),$$

with the inequality, $\sigma(\tau) < \bar{m}(\tau)$, holding provided $\tau \neq \tau_0, \tau_1$. Hence, all shock layer curves lie below $\bar{M}^*$ in R. (The same proof shows that $\bar{M}^*$ is the maximum among all monotonic arcs dominated by M^*.) Replacing M^* by $\bar{M}^*$, we find that the same argument as in the case that M^* is monotonic proves that, for sufficiently small μ/λ, the shock layer curves $S(\lambda, \mu)$ lie in any preassigned neighborhood of $\bar{M}^*$. Thus we can state,

---

[8] That $Z_0 \in \bar{M}^*$ is clear; for, if not, there is a $\bar{\tau}$, $\tau_1 < \bar{\tau} < \tau_0$, for which $l(\bar{\tau}) < m(\bar{\tau}) < m(\tau_0) = l(\tau_0)$, contradicting the monotonic decreasing character of $l(\tau)$.

LEMMA 2. *Let $\bar{M}^*$ be the arc defined by the function,*

$$\theta = \bar{m}(\tau) = \operatorname*{Min}_{\tau_1 \leq t \leq \tau} m(t), \qquad (\tau_1 \leq \tau \leq \tau_0);$$

and let G be any open neighborhood of $\bar{M}^*$. Then all shock layer curves $S(\lambda, \mu)$, for which μ/λ is sufficiently small, lie entirely in G.

Consider now the reduced systems,

$$\text{(10a)} \qquad \begin{aligned} \mu(d\tau/dx) &= M(\tau, \theta) \\ 0 &= L(\tau, \theta) \end{aligned}; \qquad \text{(10b)} \qquad \begin{aligned} \lambda(d\theta/dx) &= L(\tau, \theta) \\ 0 &= M(\tau, \theta). \end{aligned}$$

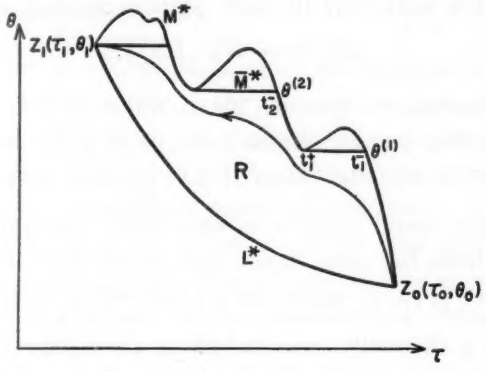

FIG. 2.

The former, (10a), are the equations of a viscous, thermally non-conducting flow, and (10b) of a non-viscous, thermally conducting flow. Let us call any solution of the reduced system (10a), for τ in $[\tau_1, \tau_0]$ and for all x, a *thermally non-conducting shock layer*; (the constant solutions, $(\tau(x), \theta(x)) = Z_0, Z_1$, are thus included in this definition, chiefly for convenience). Then we have,

THEOREM 3. *Let $\bar{S}(x; \bar{\mu})$ be a thermally non-conducting shock layer with viscosity $\bar{\mu}$. Then, the family of shock layers, $S(x; \lambda, \bar{\mu})$, if suitably parametrized, approaches $\bar{S}(x; \bar{\mu})$ as $\lambda \to 0$; that is,*

$$\lim_{\lambda \to 0} S(x; \lambda, \bar{\mu}) = \bar{S}(x; \bar{\mu}),$$

the convergence being uniform in x in every finite interval.

Proof. Let $\zeta = \bar{S}(\bar{x}; \bar{\mu}) = (\bar{\tau}(\bar{x}; \bar{\mu}), \bar{\theta}(\bar{x}; \bar{\mu}))$ be a point on L^*, and x_λ, values such that $x_\lambda \to \bar{x}$ as $\lambda \to 0$. Choose the parametrization of the shock layers, $S(x; \lambda, \bar{\mu}) = (\tau(x; \lambda, \bar{\mu}), \theta(x; \lambda, \bar{\mu}))$, so that $\lim_{\lambda \to 0} S(x_\lambda; \lambda, \bar{\mu}) = \zeta$; this

is possible by virtue of Lemma 1. Also from this lemma, we have for every shock layer,

$$\theta = l(\tau) + \epsilon(\tau, \lambda), \tag{11}$$

where $\lim_{\lambda \to 0} \epsilon(\tau, \lambda) = 0$ uniformly in $\tau_1 \leq \tau \leq \tau_0$, and therefore in $-\infty < x < +\infty$. Thus,

$$\mu d\tau(x; \lambda, \mu)/dx = M(\tau, l(\tau) + \epsilon(\tau, \lambda)).$$

Because of the choice of the $\tau(x; \lambda, \mu)$, we have from this equation, that $\lim_{\lambda \to 0} \tau(x; \lambda, \mu)$ exists uniformly in every finite x-interval and satisfies

$$\mu d\tau/dx = M(\tau, l(\tau)),$$

that is, satisfies the reduced system (10a). Since there is a unique solution of the reduced system passing through the point ζ for $x = \bar{x}$, we conclude $\lim_{\lambda \to 0} \tau(x; \lambda, \mu) = \bar{\tau}(x; \mu)$, and from (11) $\lim_{\lambda \to 0} \theta(x; \lambda, \mu) = \bar{\theta}(x; \mu)$, these limits, by the above, being uniform in every finite interval.

We observe from the above proof that if the shock layers are so parametrized that $\lim_{\lambda \to 0} S(x; \lambda, \mu)$ exists for a single value of x, then it exists for all x and defines a thermally non-conducting shock layer.

For λ fixed and $\mu \to 0$ the limit solution of the reduced system is no longer continuous in general, as the following Theorem 4 shows. This points up a basic difference in the effects of viscosity and heat conduction on the structure of the shock layers.

To simplify considerations, it will be assumed in the following that M^* has only a finite number of minima, so that the arc $\bar{M}^*$ contains at most a finite number of intervals on which θ is constant. We note that Z_0 cannot lie in such an interval, while Z_1 may or may not, in general. In particular, for an ideal gas, M^* is a parabola; if it is not monotonic, $\bar{M}^*$ consists of the segment of $\theta = \theta_1$ intercepted by M^* plus the arc of M^* joining the segment to Z_0.

Let the function $\theta = \bar{m}(\tau)$, $(\tau_1 \leq \tau \leq \tau_0)$, be constant on the intervals, $[t_i^+, t_i^-]$, $(i = 1, \cdots, n)$, which are ordered so that $t_i^- > t_i^+ > t_{i+1}^-$, (Fig. 2). We enlarge the notion of solution of the reduced system (10b) by admitting certain *discontinuous* solutions, namely, functions $\bar{\tau}(x)$, $\bar{\theta}(x)$ satisfying (10b) for $\bar{\tau}$ in $[\tau_1, \tau_0]$, except at points $x = x_i$, $i = 1, \cdots, n$, where $\bar{\tau}(x_i - 0) = t_i^-$, $\bar{\tau}(x_i + 0) = t_i^+$; these solutions are uniquely determined in $-\infty < x < +\infty$ up to an x-translation. If $n = 0$, that is, if $m(t)$ is strictly monotonic, then the solutions are, of course, continuous. Any solution of (10b) for τ in

$[\tau_1, \tau_0]$ and for all x, whether continuous, or discontinuous in the above sense, we term a *non-viscous shock layer*. We now prove,

THEOREM 4. *Let $\bar{S}(x;\bar{\lambda})$ be a non-viscous shock layer with heat conductivity $\bar{\lambda}$. Then, the family of shock layers, $S(x;\bar{\lambda},\mu)$, if suitably parametrized, approaches $\bar{S}(x;\bar{\lambda})$ as $\mu \to 0$; that is,*

$$\lim_{\mu\to 0} S(x;\bar{\lambda},\mu) = \bar{S}(x;\bar{\lambda}),$$

the convergence being uniform in x in any closed interval not containing a point of discontinuity of $\bar{S}(x;\bar{\lambda})$.

Proof. If M^* is strictly monotonic, the theorem is proved exactly as Theorem 3. If, however, this is not the case, we proceed as follows. Let $\bar{S}(x;\bar{\lambda}) \equiv \bar{S}(x) = (\bar{\tau}(x), \bar{\theta}(x))$; (for notational convenience in the following, we omit reference to $\bar{\lambda}$, since it will be the common value of the heat conductivity for all the shock layers under consideration). Let $\zeta = \bar{\tau}(\bar{x}), \bar{\theta}(\bar{x}))$ be a point on M^*, $(\bar{x} < \bar{x}_1)$, where $\bar{x}_1$ is the first point of discontinuity of $\bar{\tau}(x)$. For any set x_μ, such that $x_\mu \to \bar{x}$ as $\mu \to 0$, let the shock layers $S(x;\mu) = (\tau(x;\mu), \theta(x;\mu))$ be so parametrized that $\lim_{\mu\to 0} S(x_\mu;\mu) = \zeta$. We have, by Lemma 2,

$$\tau = m^{-1}(\theta) + \epsilon(\theta,\mu), \tag{12}$$

where $m^{-1}(\theta)$ is the inverse of $m(\tau)$ for $t_1^- \leq \tau \leq \tau_0$, $(\theta_0 \leq \theta \leq \theta^{(1)} = m(t_1^-))$, and $\lim_{\mu\to 0} \epsilon(\theta,\mu) = 0$ uniformly in every closed interval of $\theta_0 \leq \theta < \theta^{(1)}$. For $\theta(x;\mu)$ we therefore have,

$$\bar{\lambda} d\theta(x;\mu)/dx = L[m^{-1}(\theta) + \epsilon(\theta,\mu), \theta]. \tag{13}$$

Since $\theta(x_\mu,\mu) \to \theta(\bar{x})$ and $x_\mu \to \bar{x}$ as $\mu \to 0$, and $\bar{\tau} = m^{-1}(\bar{\theta})$ for $\theta_0 \leq \theta < \theta^{(1)}$, if follows that $\lim_{\mu\to 0} \theta(x;\mu) = \bar{\theta}(x)$, $-\infty < x < \bar{x}_1$, and from (12), $\lim_{\mu\to 0} \tau(x;\mu) = \bar{\tau}(x)$, the convergence in both cases being uniform in every closed interval to the left of $x = \bar{x}_1$. To prove convergence for $x > \bar{x}_1$, consider first the case that $\tau_1 \neq t_1^+$. By the preceding there are values $x_\mu^{(1)} < \bar{x}_1$, such that $x_\mu^{(1)} \to \bar{x}_1$, $\tau(x_\mu^{(1)}, \mu) \to t_1^-$, and $\theta(x_\mu^{(1)}, \mu \to \theta^{(1)}$, as $\mu \to 0$. With the same parametrization of the $S(x;\mu)$ as chosen above, let ξ_μ be values for which $\theta(\xi_\mu;\mu) = \theta^{(1)}$; then $\tau(\xi_\mu;\mu) \to t_1^+$ as $\mu \to 0$ (Lemma 2). We show now that $|\xi_\mu - x_\mu^{(1)}| \to 0$ with μ. For, $L(\tau,\theta)$ is bounded away from zero in the neighborhood of the segment $\theta = \theta^{(1)}$ of $\bar{M}^*$; say $L(\tau,\theta) > k > 0$ in such a neighborhood; and, if μ is sufficiently small, the arcs of $S(x;\mu)$ for

$x_\mu^{(1)} \leq x \leq \xi_\mu$ lie in this neighborhood (Lemma 2). For these μ, we have from the first member of (E),

$$(14) \quad |\xi_\mu - x_\mu^{(1)}| \leq (\bar{\lambda}/k)[\theta(\xi_\mu, \mu) - \theta(x_\mu^{(1)}, \mu)] = (\bar{\lambda}/k)[\theta^{(1)} - \theta(x_\mu^{(1)}, \mu)],$$

which approaches zero as $\mu \to 0$. Hence $\xi_\mu \to \bar{x}_1$. We consider equation (12) and (13) again, where now $m^{-1}(\theta)$ is the inverse of $\theta = m(\tau)$ for $t_2^- \leq \tau \leq t_1^+$, $(\theta^{(1)} \leq \theta \leq \theta^{(2)} = m(t_2^-))$, and $\lim_{\mu \to 0} \epsilon(\theta, \mu) = 0$ uniformly in every closed interval of $\theta^{(1)} \leq \theta < \theta^{(2)}$. As before, since $\theta(\xi_\mu, \mu) \to \theta^{(1)} = \bar{\theta}(x_1)$ as $\xi_\mu \to \bar{x}_1$, and $\bar{\tau} = m^{-1}(\bar{\theta})$ for $\theta^{(1)} \leq \bar{\theta} \leq \theta^{(2)}$, and also since the solution of

$$\bar{\lambda} d\theta/dx = L(m^{-1}(\theta), \theta)$$

is unique for $\theta(\bar{x}_1) = \theta^{(1)}$, and $\theta^{(1)} \leq \theta \leq \theta^{(2)}$, it follows that $\lim_{\mu \to 0} \theta(x; \mu) = \bar{\theta}(x)$ and $\lim_{\mu \to 0} \tau(x; \mu) = \bar{\tau}(x)$ for $\bar{x} \leq x < \bar{x}_2$, the convergence being uniform in the closed intervals of $(\bar{x}_1, \bar{x}_2)$. One proceeds in this way until the arc of $\bar{M}^*$ containing Z_1 is reached. If Z_1 is not contained in one of the intervals of $\bar{M}^*$, then the above process shows the uniform convergence of $S(x; \mu)$ to $(\bar{\tau}(x), \bar{\theta}(x))$ in every closed subinterval of the half line $\bar{x}_n < x < \infty$. If, however, Z_1 belongs to one of the intervals of $\bar{M}^*$, namely $\theta = \theta_1$, then, by definition of the discontinuous non-viscous shock layer, $\tau_1 = t_n^+$, and $(\bar{\tau}(x), \bar{\theta}(x)) = (\tau_1, \theta_1) = Z_1$ for all x in $\bar{x}_n < x < \infty$. In this case, we prove the convergence, $S(x; \mu) \to Z_1$ for $\bar{x}_n < x < \infty$, as follows. For any small $\delta > 0$, let $R(\delta)$ be the subregion of R outside the circle of radius δ about Z_1. Let $k(\delta)$ be such that $L(\tau, \theta) \geq k(\delta) > 0$ in a neighborhood of $\theta = \theta_1$ in $R(\delta)$. If now the values $x_\mu^{(n)} < \bar{x}_n$ are such that $x_\mu^{(n)} \to \bar{x}_n$, $\tau(x_\mu^{(n)}, \mu) \to t_n^-$, and $\theta(x_\mu^{(n)}, \mu) \to \theta^{(n)} = \theta_1$, as $\mu \to 0$, then, for any arguments $\xi_\mu > x_\mu^{(n)}$ such that $S(\xi_\mu; \mu) \in R(\delta)$, it follows from the first of equations (E),

$$|\xi_\mu - x_\mu^{(n)}| \leqq (\bar{\lambda}/k(\delta))[\theta_1 - \theta(x_\mu^{(n)}; \mu] \to 0, \text{ as } \mu \to 0;$$

in other words, those $x > \bar{x}_n$ for which $\tau(x; \mu) - \tau_1 > \delta$ lie in an interval about $\bar{x}_n$ which grows arbitrarily small as $\mu \to 0$. Since δ is arbitrary, this proves the uniform convergence, $\tau(x; \mu) \to \tau_1, \theta(x; \mu) \to \theta_1$, in any closed half line of $\bar{x}_n < x < \infty$, and completes the proof of the theorem.

One observes from the above proof that if the shock layers are so parametrized that $\lim_{\mu \to 0} S(x; \bar{\lambda}, \mu)$ exists for a single value of x, then it exists for all x and defines a non-viscous shock layer.

To complete discussion of the limit behavior of the shock layers, it

remains to establish the existence of the iterated limits as $\lambda, \mu \to 0$, and to show their equality with the double limit (and therefore with a shock wave). We require an appropriate parametrization of the shock layers, with respect to which the double limit and the two iterated limits exist simultaneously. For this, take a circle of sufficiently small radius ϵ about Z_0 (it suffices $\epsilon < \tau_0 - t_1^-$); this intersects each shock layer curve, $S(\lambda, \mu)$, at exactly one point, which we designate by $\zeta_{\lambda\mu}$; and let ζ_λ, ζ_μ represent the point of intersection with M^* and L^*, respectively. We note that $\lim_{\lambda\to 0} \zeta_{\lambda\mu} = \zeta_\mu$ (Lemma 1), and $\lim_{\mu\to 0} \zeta_{\lambda\mu} = \zeta_\lambda$ (Lemma 2). For fixed ξ, now let

$$x_{\lambda\mu} = \xi - \tfrac{1}{2}\delta_{\lambda\mu}, \qquad \bar{x}_\mu = \xi - \tfrac{1}{2}\delta_\mu, \qquad \bar{x}_\lambda = \xi - \tfrac{1}{2}\delta_\lambda,$$

where,

$$\delta_{\lambda\mu} = (\eta/C(\epsilon))[\theta_1 - \theta_1 + \tau_0 - \tau_1], \qquad \eta = \mathrm{Max}(\lambda, \mu),$$

$$\delta_\mu = (\mathrm{Max}\,\mu/C(\epsilon))\,[\theta_1 - \theta_0 + \tau_0 - \tau_1],$$

$$\delta_\lambda = (\mathrm{Max}\,\lambda/C(\epsilon))[\theta_1 - \theta_0 + \tau_0 - \tau_1],$$

$C(\epsilon)$ being the same bound as in inequality (7) of Theorem 2. We make the assignment $S(x_{\lambda\mu}; \lambda, \mu) = \zeta_{\lambda\mu}$; this will prove to be the desired parametrization. From the proof of Theorem 2, we see that, since ξ is contained in all the intervals $(x_{\lambda\mu}, x_{\lambda\mu} + \delta_{\lambda\mu})$ as $\eta \to 0$, it must follow,

$$\lim_{\substack{\lambda\to 0 \\ \mu\to 0}} S(x; \lambda, \mu) = Z_0, \quad -\infty < x < \xi$$
$$= Z_1, \quad \xi < x < +\infty.$$

Also, since for fixed μ,

$$\lim_{\lambda\to 0} \delta_{\lambda\mu} = (\mathrm{Max}\,\mu/C(\epsilon))[\theta_1 - \theta_0 + \tau_0 - \tau_1] = \delta_\mu,$$

we have $\lim_{\lambda\to 0} x_{\lambda\mu} = \xi - \tfrac{1}{2}\delta_\mu = \bar{x}_\mu$, $\lim_{\lambda\to 0} S(x_{\lambda\mu}; \lambda, \mu) = \zeta_\mu$. Therefore as in the proof of Theorem 3, $\lim_{\lambda\to 0} S(x; \lambda, \mu) = (\bar{\tau}(x; \mu), \theta(x; \mu))$ where $(\bar{\tau}(x; \mu), \bar{\theta}(x; \mu))$ is the solution of the reduced system (10a) which satisfies $(\bar{\tau}(\bar{x}_\mu; \mu), \bar{\theta}(\bar{x}_\mu; \mu)) = \zeta_\mu$. This, combined with the existence of the double limit, establishes the existence of the iterated limit, $\lim_{\mu\to 0} (\lim_{\lambda\to 0} S(x; \lambda, \mu))$, and its equality with the double limit. Identical considerations apply to $\lim_{\lambda\to 0} (\lim_{\mu\to 0} S(x; \lambda, \mu)$. Consequently, we may state,

THEOREM 5. *Let $Z_0 = (\tau_0, \theta_0)$, $Z_1 = (\tau_1, \theta_1)$ be initial and final states of the shock wave, $Z = Z_0$, $-\infty < x < \xi$, $Z = Z_1$, $\xi < x < +\infty$. Then the shock layers $S(x; \lambda, \mu)$ can be parametrized so that,*

$$\lim_{\substack{\lambda \to 0 \\ \mu \to 0}} S(x; \lambda, \mu) = \lim_{\lambda \to 0} \left(\lim_{\mu \to 0} S(x; \lambda, \mu) \right)$$

$$= \lim_{\mu \to 0} \left(\lim_{\lambda \to 0} S(x; \lambda, \mu) \right) = \begin{matrix} Z_0, & -\infty < x < \xi, \\ Z_1, & \xi < x < +\infty. \end{matrix}$$

It is clear, conversely, that the iterated limit, if it exists (for some parametrization), must be either a constant or a shock wave.

In the case of polytropic gases with λ, μ taken independent of τ and θ, the convergence of the iterated limits to a shock wave is visible from explicit formulas for the thermally non-conducting shock layers, $\lim_{\lambda \to 0} S(x; \lambda, \mu)$, and for the non-viscous shock layers, $\lim_{\mu \to 0} S(x; \lambda, \mu)$, (presupposing, however, the existence of these single limits; see Becker [2], for example).

The advantages of the (τ, θ) plane as phase plane should be noted. In these variables, the right members of equations (E), and therefore the curves L and M, are independent of λ and μ. This fact makes the proof of Theorem 1 quite simple, and also permits study, with respect to a fixed geometry, of the dependence of the shock layers on λ and μ. In contrast, the topologies of other phase planes, such as the (τ, p) plane used by Weyl in [1], are such as to obscure a satisfactory proof of the existence and limit theorems.

3. Proof of Conditions (A) — (D). It remains to show that conditions (A)-(D) of Section 1 are satisfied for the fluids under consideration. We examine first the polytropic gases because of their particular importance and simplicity. For polytropic gases, $e(\tau, \theta) = \alpha\theta$, $p(\tau, \theta) = \beta\theta/\tau$, where $\alpha, \beta > 0$ are constants of the medium. Thus

$$L(\tau, \theta) = b[\alpha\theta - b^2/2(\tau - a)^2 - c], \quad M(\tau, \theta) = 1/b[\beta\theta/\tau + b^2(\tau - a)].$$

The Z-plane is the quadrant $\tau > 0$, $\theta > 0$, and the constants a, b, c are assumed to be such that the parabolas L: $L(\tau, \theta) = 0$, and M: $M(\tau, \theta) = 0$, intersect in two points, $Z_0 = (\tau_0, \theta_0)$, $Z_1 = (\tau_1, \theta_1)$, $\tau_0 > \tau_1$. Conditions (A)-(D) are now easily seen to hold. (A) $M_\theta, L_\tau > 0$, since $\alpha, \beta, b > 0$; (B) the parabolas L, M have only one component in the Z-plane, and cannot intersect in more than the two assumed points, Z_0, Z_1; (C) $L_\tau = -b^2/2(\tau - a) > 0$ for $\tau < a$, hence for $\tau_1 \leq \tau \leq \tau_0 < a$; (D) follows from the fact that M lies above L between their points of intersection.

We turn now to the general class of fluids defined by Weyl. These are characterized by the following conditions:

I. $d\tau/dp)_{S=\text{const.}} < 0$, $S =$ entropy;

Ia. $S_p(\tau, p) > 0$, $\theta_p(\tau, p) > 0$;

II. $d^2\tau/dp^2)_{S=\text{const.}} > 0$;

III. in the continuous process of adiabatic compression one can raise pressure arbitrarily high;

IV. the thermodynamic state Z is uniquely specified by pressure p and specific volume τ, and the points (τ, p) representing the possible states Z in a (τ, p) diagram form a convex region.

To these conditions must of course be added the basic thermodynamic relation,

V. $de = \theta dS - p d\tau$.

This class of fluids evidently includes the polytropic gases, and, as Weyl has shown, the theory of shock waves generalizes in entirety to these fluids.

For fluids subject to the above conditions, Weyl proves a number of important results, among which we need the following.

(a) The Hugoniot contour,

$$H(Z, Z_0) = e(\tau, p) - e(\tau_0, p_0) - \tfrac{1}{2}(p + p_0)(\tau_0 - \tau) = 0$$

is a simple curve on which $s = (p - p_0)/(\tau_0 - \tau)$ grows strictly monotonically from $m_0 = -dp/d\tau)_{S=S_0}$ at Z_0 to $+\infty$ as Z moves on the upper branch, $p > p_0$, $\tau < \tau_0$.

(b) If $H(Z_1, Z_0) = 0$, Z_1 lying on the upper branch of $H(Z, Z_0) = 0$, then the straight line $(p - p_0)/(\tau_0 - \tau) = (p_1 - p_0)/(\tau_0 - \tau_1) = b^2$ intersects the adiabatic $S = S_0$ at a point $Z = Z_A$, Z_1 lying between Z_0 and Z_A, and $m_0 < b^2 < m_1 = -dp/d\tau(Z_1)\big|_{S=S_1}$.

To prove conditions (A)-(D) of Section 1, consider first the mapping $(\tau, p) \to (\tau, \theta)$ defined by $\theta = \theta(\tau, p)$, $\tau = \tau$. This mapping is 1-1 as well as continuous, for, if both (τ_1, p_1) and (τ_2, p_2) map into the same point (τ, θ), then $\tau_1 = \tau = \tau_2$, and, since $\theta_p > 0$ in the convex (τ, p) plane, it follows also that $p_1 = p_2$. Consequently, the mapping is topological and the $Z(\tau, \theta)$ plane is simply connected.

In the following, let $Z_0(\tau_0, p_0)$, $Z_1(\tau_1, p_1)$, be points for which $H(Z_1, Z_0) = 0$, with $p_1 > p_0$, $\tau_1 < \tau_0$; (either $p_1 > p_0$, $\tau_1 < \tau_0$, or $p_1 < p_0$, $\tau_1 > \tau_0$). Let

$$bM(\tau, p) \equiv p + b^2(\tau - a), \qquad (1/b)L(\tau, p) \equiv e(\tau, p) - \tfrac{1}{2}b^2(\tau - a)^2 - c,$$

$b > 0$, a, c being determined by relations (4), (5). The functions $M(\tau, \theta)$, $L(\tau, \theta)$ are defined in the obvious way:

$$M(\tau, \theta) \equiv M(\tau, p(\tau, \theta)), \qquad L(\tau, \theta) \equiv L(\tau, p(\tau, \theta)).$$

Condition (A) is easily proved, for

$$bM_\theta(\tau, \theta) = p_\theta(\tau, \theta) > 0, \qquad \text{(by Ia)},$$

and

$$(1/b)L_\theta(\tau, \theta) = e_\theta(\tau, \theta) = \theta S_\theta(\tau, \theta), \qquad \text{(by V)}$$
$$= S_p(\tau, p)p_\theta(\tau, \theta) > 0,$$

so that $L_\theta(\tau, \theta) > 0$.

To prove (B), we observe from (a) that $Z_0(\tau_0, p_0)$, $Z_1(\tau_1, p_1)$ are the only simultaneous solutions of $L(\tau, p) = 0$, $M(\tau, p) = 0$ (Weyl's Theorem 7), and therefore $Z_0(\tau_0, \theta_0)$, $Z_1(\tau_1, \theta_1)$ are the only solutions of $L(\tau, \theta) = 0$, $M(\tau, \theta) = 0$. Also, $M(\tau, p) = 0$ only on the straight line M joining Z_0 and Z_1, and since the (τ, p) plane is convex, M consists of only one component. The set in the (τ, θ) plane on which $M(\tau, \theta) = 0$ must therefore also consist of a single curve, namely, the image of M. Letting $\mathfrak{L}$ designate the set of points (τ, p) on which $L(\tau, p) = 0$, it remains only to show that the component $L \subset \mathfrak{L}$, which passes through Z_1 also passes through Z_0. Since

$$(1/b)L_p(\tau, p) = e_p(\tau, p) = \theta S_p(\tau, p) > 0,$$

the curve L is representable as a single valued function $p = l(\tau)$, with slope,

$$(15) \qquad dl(\tau)/d\tau = -(e_\tau(\tau, p) - b^2(\tau - a))/e_p(\tau, p)$$
$$= -(\theta S_\tau(\tau, b) - p - b^2(\tau - a))/\theta S_p(\tau, p), \quad ((\tau, p) \in L).$$

In particular, at Z_1, L has the slope,

$$-S_\tau/S_p(\tau_1, p_1) = dp/d\tau(Z_1)\big|_{S=S_1} = -m_1 < -b^2.$$

Consequently, L, for increasing τ, enters the convex region D contained between the adiabatic $S = S_0$ and the straight line M (Fig. 3). Since D

lies below M, $p + b^2(\tau - a) < 0$ in D, so that, by (15), $l(\tau)$ is monotonically decreasing function of τ in this region. The curve L cannot terminate in D, must therefore intersect either M, which can only occur at Z_0, or the adiabatic, $S = S_0$.

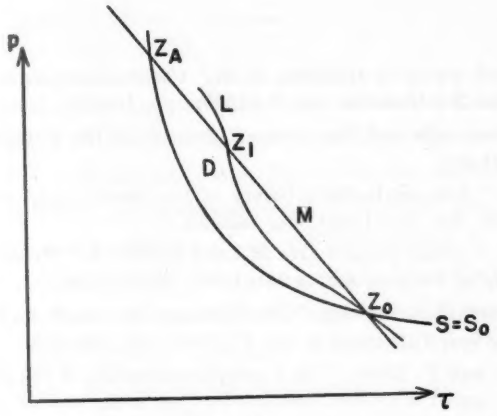

FIG. 3.

But on $S = S_0$,

$$(1/b)\,(dL(\tau, p)/d\tau)\Big)_{S=S_0} = de/d\tau\Big)_{S=S_0} - b^2(\tau - a)$$
$$= [-p - b^2(\tau - a)]_{S=S_0} > 0, \text{ for } \tau_A < \tau < \tau_0;$$

hence $L(\tau, p) \neq 0$ on $S = S_0$ between Z_0 and Z_1. L must therefore pass through Z_0 as well as Z_1. This proves that condition (B) is satisfied.

For condition (C),

$$(1/b)\,L_\tau(\tau, \theta) = e_\tau(\tau, \theta) - b^2(\tau - a) = \theta S_\tau(\tau, \theta) - p - b^2(\tau - a);$$

but from V, we have, $S_\tau(\tau, \theta) = p_\theta(\tau, \theta) > 0$,

$$\therefore (1/b)\,L_\tau(\tau, \theta) = \theta p_\theta(\tau, \theta) - p - b^2(\tau - a) > 0 \text{ on } L, \text{ for } \tau_1 \leqq \tau \leqq \tau_0.$$

(D) follows from the relations,

$$M_\tau/M_\theta(\tau, \theta) - L_\tau/L_\theta(\tau, \theta) = [M_\tau/M_p(\tau, p) - L_\tau/L_p(\tau, p)]\theta_p(\tau, p)$$
$$= (b^2 - m_0)\theta_p(Z_0) > 0 \quad \text{at } Z_0$$
$$= (b^2 - m_1)\theta_p(Z_1) < 0 \quad \text{at } Z_1.$$

This completes the proof that a Weyl fluid satisfies the conditions (A)-(D) which were assumed in obtaining Theorems 1-5.

BIBLIOGRAPHY.

[1] H. Weyl, "Shock waves in arbitrary fluids," *Communications on Pure and Applied Applied Mathematics*, vol. 2 (1949), pp. 103-122.

[2] R. Becker, "Stosswelle und Detonation," *Zeitschrift für Physik*, vol. 8 (1921-22), pp. 321-362.

[3] L. H. Thomas, "Note on Becker's theory of the shock front," *Journal of Chemical Physics*, vol. 12 (1944), pp. 449-452.

[4] Lord Rayleigh, "Aerial plane waves of finite amplitude," *Proceedings of the Royal Society of London*, Ser. A 84 (1910), pp. 247-284.

[5] A. E. Puckett and H. J. Stewart, "The thickness of a shock wave in air," *Quarterly of Applied Mathematics*, vol. 7 (1950), pp. 457-463.

[6] M. Morduchow and P. Libby, "On a complete solution of the one-dimensional flow equations of a viscous, heat-conducting, compressible gas," *Journal of the Aeronautical Sciences*, vol. 16 (1949), pp. 674-684, 704.

[7] R. Courant and K. Friedrichs, *Supersonic flow and shock waves*, Interscience, New York (1948).

[8] G. I. Taylor and J. W. Maccoll, *The Mechanics of compressible fluids*, vol. III of W. F. Durand, *Aerodynamic Theory*, Springer, Berlin (1934), pp. 218-222.

[9] P. Duhem, *Recherches sur l'hydrodynamique*, Gauthier-Villars, Paris (1903); Première partie, Ch. I. = Annales de la Faculté des Sciences Toulouse, (2) vol. 3 (1901), pp. 315-377.

EQUIVALENCES OF MEASURE SPACES.*

By I. E. SEGAL.

1. Introduction. For any measure space we define the "measure ring" as the Boolean ring of all measurable sets (including not necessarily σ-finite sets) modulo the ideal of all null sets, the "σ-finite measure ring" as the subring of elements which are residue-classes of σ-finite sets, and the "multiplication algebra" as the algebra of all multiplications of elements of the Hilbert space of square integrable functions over the space by bounded measurable functions. Two measure spaces have algebraically isomorphic measure rings (we say then that they are "strongly equivalent") if and only if their multiplication algebras are unitarily equivalent, or alternatively, if their multiplication algebras are algebraically isomorphic. This is of special interest in view of the fact [5] that every maximal abelian self-adjoint algebra of bounded operators on a Hilbert space is unitarily equivalent to the multiplication algebra of some measure space. The multiplication algebra of a measure space is maximal abelian if and only if its measure ring is complete (as a partially ordered set), or alternatively, if and only if it is strongly equivalent to a direct sum (see below) of finite measure spaces. The class of measure spaces with these properties (we call such spaces "localizable") constitutes in some ways a more natural generalization of the σ-finite measure spaces, than the class of arbitrary measure spaces. In particular, for a measure space to be localizable is equivalent to the validity for the space of the conclusion of the Radon-Nikodym theorem, or alternatively to the validity of the conclusion of the Riesz representation theorem for continuous linear functionals on the Banach space of integrable functions. Every measure space is metrically equivalent (by which we mean that there is a measure-preserving isomorphism between the σ-finite measure rings—roughly speaking this means the spaces are equivalent as far as integration over them is concerned) to a localizable space, and this latter space is essentially unique.

More specifically, every measure space is metrically equivalent to a unique perfect measure space, this being a regular locally compact space in which for every measurable set of finite measure there is a unique continuous function on the space equal nearly everywhere to the characteristic

* Received June 25, 1950.

function of the set; and any metric equivalence between localizable spaces is uniquely extendable to a strong equivalence between them. These perfect measure spaces, which are a development of spaces treated by Stone, Freudenthal, and especially Kakutani [3] (the metric equivalence of a finite measure space with a perfect space being proved here), are useful in eliminating certain types of measurability problems (cf. [4]) for they have many of the properties of discrete spaces.

The present paper arose from an investigation of the decomposition and multiplicity theory of algebras of operators on Hilbert spaces, and as the results are of some independent interest they are being published separately. Among other side results we mention the fact that any uniformly locally compact regular measure space, in particular any locally compact group under Haar measure, is localizable. Any two measure spaces have algebraically isomorphic σ-finite measure rings if and only if the corresponding Lebesgue spaces over them are isomorphic as partially ordered linear spaces, and are metrically equivalent if and only if there is a more restricted type of isomorphism between certain function spaces over the measure spaces.

2. Technical preliminaries. In this section we fix the notation and terminology for the remainder of the paper and prove some simple results which are basic in the following.

DEFINITIONS 2.1. A *conditional σ-ring* $\mathfrak{R}$ of subsets of a set R is a collection of subsets which is closed under relative complementation and union, and has the further property that if $\{E_n\}$ is a sequence of elements in R such that $E_n \subset E$ for all n, where $E \in R$, then $\bigcup_n E_n \in \mathfrak{R}$. A *measure space* $(R, \mathfrak{R}\ r)$ is the system composed of a set R, a conditional σ-ring $\mathfrak{R}$ of subsets of R, and a countably-additive, finite (non-negative) measure r defined on $\mathfrak{R}$, such that if $\{E_n\}$ is a sequence of mutually disjoint sets in $\mathfrak{R}$ for which the series $\Sigma_n r(E_n)$ is convergent, then $\bigcup_n E_n \in \mathfrak{R}$.

Remark 2.1. The term "measure space" is commonly used in different senses, e. g. frequently in reference to a system $(S, \mathfrak{S}, s)$ composed of a set S, a σ-ring $\mathfrak{S}$ of subsets of S (i. e. a conditional σ-ring which is closed under arbitrary countable unions), and a not-necessarily-finite, countably additive measure s on $\mathfrak{S}$. Any such measure space becomes a measure space in the sense of Definition 2.1 when its σ-ring $\mathfrak{S}$ of sets is replaced by the subring $\mathfrak{S}_0$ of sets in $\mathfrak{S}$ of finite measure, and this replacement affects neither the (suitably-defined) integrability of functions on the space, nor the values of the integrals of integrable functions. It is advantageous, for formal and

notational purposes, to replace $\mathcal{S}$ by $\mathcal{S}_0$ because the elements of $\mathcal{S}$-$\mathcal{S}_0$ must generally be treated in terms of the elements of $\mathcal{S}_0$. We mention also that to replace $\mathcal{S}$ by the σ-ring of σ-finite elements of $\mathcal{S}$ (i. e. elements which are countable unions of elements of $\mathcal{S}_0$) is less advantageous, mainly because the distinction between σ-finite and arbitrary elements of $\mathcal{S}$-$\mathcal{S}_0$ is often an artificial one, from the viewpoint of the present paper.

DEFINITIONS 2. 2. *Sets in measure spaces.* A set K in a measure space $M = (R, \mathcal{R}, r)$ (i. e. a subset of R) is called *measurable* if $K \cap E \in \mathcal{R}$ for all $E \in \mathcal{R}$; a *null set* if $K \cap E$ is an element of $\mathcal{R}$ on which r vanishes, for all $E \in \mathcal{R}$. The *measure of a measurable set* K in M is defined as the least upper bound (finite or ∞) of $r(E)$, as E varies over the subsets of K which are in $\mathcal{R}$; and is denoted by $r(K)$. A measurable set of finite measure is called a *chunk*. A statement which is a function on a measure space is said to be true *nearly everywhere* (abbreviated "n. e.") if it is valid at all points of the space, except for a null set. Two subsets K and L of a measure space are said to be *equivalent* if their symmetric difference $(K\text{-}L) \cup (L\text{-}K)$, which will be denoted as $K \ominus L$, is a null set. A set K in M is called *σ-finite* if it is contained in a countable union of chunks. M is called *σ-finite* if R is *σ-finite* and *finite* in case R is itself a chunk.

THEOREM 2. 1. *The collection of all measurable subsets of a measure space is a complemented σ-ring on which the measure is countably additive.*

The proof of this uses standard methods in a straightforward fashion, and so we omit details.

DEFINITIONS 2. 3. *Functions on measure spaces.* A complex-valued function on a measure space is said to be *measurable* if it is the pointwise limit of a sequence of finite linear combinations of characteristic functions of measurable sets. Such a function is said to *vanish at infinity* on the space if for every positive number e, the set where the function exceeds e in absolute value is contained in a chunk. A complex valued function f on a measure space $M = (R, \mathcal{R}, r)$ is said to be *integrable* if there exists a sequence f_n of finite linear combinations of characteristic functions of elements of $\mathcal{R}$ which converges pointwise to f and which is such that $\int |f_n(p) - f_m(p)|\, dr(p)$ converges to 0 as $m, n \to \infty$, where for any finite linear combination $h = \Sigma \alpha_i \chi_{E_i}$ of characteristic functions of mutually disjoint elements E_i of $\mathcal{R}$ (for any set S, χ_S denotes the characteristic function of S), $\int h(p)\, dr(p)$ is defined as $\Sigma \alpha_i r(E_i)$; the value of $\lim_n \int f_n(p)\, dr(p)$ is then called the value of the integral of f, and denoted as $\int f(x)\, dr(x)$. A complex-valued function f on

M is said to be α-th-power integrable (or square-integrable, in case $\alpha = 2$), where $\alpha \geq 1$, if it is measurable and $|f(x)|^\alpha$ is integrable. For any such function the value of $\{\int |f(x)|^\alpha dr(x)\}^{1/\alpha}$ is called the *norm* of f in L_α, and denoted by $\|f\|_\alpha$; the collection of all such functions is designated as $L_\alpha(M)$. The Banach space consisting of the quotient of the normed linear space $L_\alpha(M)$ by the subspace of functions in $L_\alpha(M)$ of zero norm is called the (Banach) space of α-th-power integrable functions on M, and designated $B_\alpha(M)$. A complex-valued function f on M is said to be *essentially bounded* if there is a real number α such that $|f(x)| \leq \alpha$ except on a null set, and the g. l. b. of all such α is called the *essential bound* of f, denoted as $\|f\|_\infty$, the collection of all such functions which are measurable being denoted as $L_\infty(M)$. The Banach space, and algebra, consisting of the quotient of the normed space and algebra $L_\infty(M)$ (multiplication being the usual pointwise product), by the subspace, and ideal consisting of elements of zero norm, is called the (Banach) *algebra of bounded measurable functions on* M, and denoted $B_\infty(M)$. Two functions on a measure space are called *equivalent* if their values coincide n. e. The (real) subspace of all real-valued functions in $B_\alpha(M)$, partially-ordered by the definition: $f \leq g$ if $f(x) \leq g(x)$ n. e., is called the *partially-ordered (Banach) space of α-th-power integrable functions on* M, and denoted $RB_\alpha(M)$.

DEFINITIONS 2. 4. *Measure rings.* A *measure ring* $(\mathfrak{M}, m)$ is the system composed of a σ-Boolean ring $\mathfrak{M}$ (i. e. a Boolean ring in which each countable subset has a l. u. b.) and a countably additive measure m on $\mathfrak{M}$, such that only the zero element has measure zero, and every element is the l. u. b. of elements of finite measure. An element of a measure ring is called *σ-finite* if it is the union (l. u. b.) of a sequence of elements of $\mathfrak{M}$ of finite measure, and a measure ring is called *σ-finite* if each of its elements is.

The *measure ring* $(\mathfrak{M}, m)$ of a measure space $M = (R, \mathfrak{R}, r)$ is defined to have as $\mathfrak{M}$ the quotient of the ring of measurable sets in M by the ideal of null sets, with the measure of any residue class defined to be the measure of any representative of the class. The natural mapping of the ring of measurable sets in M onto $\mathfrak{M}$ will be denoted by θ, and the residue class corresponding to a measurable set will also be denoted by superposing a bar: thus $\theta(K) = \bar{K}$ for any measurable set K.

Remark 2. 2. It is easy to see that the mapping θ just defined is a σ-homomorphism, i. e. for any sequence of measurable sets $\{E_n; n = 1, 2, \cdots\}$, $\theta(\bigcup_n E_n) = \bigcup_n \theta(E_n)$ and $\theta(\bigcap_n E_n) = \bigcap_n \theta(E_n)$. It follows that the measure ring of a measure space is a σ-Boolean ring. It will later be shown that every

element in $\mathfrak{M}$ is the l. u. b. of elements of finite measure, so that the measure ring of M is actually a measure ring in the sense of our earlier definition.

DEFINITIONS 2. 5. *Measure rings and equivalences.* The *σ-finite* (*or finite, resp.*) *measure ring* of a measure space is the subring of the measure ring of the space consisting of all σ-finite elements (or elements of finite measure). Two measure spaces are (*strongly* or *weakly*, resp.) equivalent if there is an algebraic isomorphism between the (Boolean rings of) their (measure rings, or σ-finite measure rings resp.); and *metrically equivalent* if there is a measure-preserving algebraic isomorphism between their finite measure rings. An isomorphism which implements such an equivalence between measure spaces is called a *strong, weak, or metric equivalence*, depending on its character. A measure space $M = (R, \mathfrak{R}, r)$ is said to be *isomorphic* to a measure space $M' = (R', \mathfrak{R}', r')$ if there is a one-to-one mapping ϕ from R onto R' such that: $E \in \mathfrak{R}$ if and only if $\phi(E) \in \mathfrak{R}'$; and for $E \in \mathfrak{R}$, $r(E) = r'(\phi(E))$. Such a mapping ϕ is called an *isomorphism* on M to M'.

DEFINITION 2. 6. A measure space is called *localizable* if its measure ring is complete (as a partially-ordered set).

DEFINITION 2. 7. *Regular and perfect spaces.* A *regular* (*locally compact*) *measure space* is the system composed of a locally compact Hausdorff space R, and a measure space $(R, \mathfrak{R}, r)$ with the properties that: a) $\mathfrak{R}$ contains all compact subsets of R (so that r is finite on such sets); b) for any set E in $\mathfrak{R}$ there is an open set G in $\mathfrak{R}$ containing E and $r(E) = \text{g. l. b.}[E \subset H, H \text{ open}]r(H)$ (we frequently denote the set over which a bound is taken by placing it between the symbol for the bound and the function whose bound is indicated); c) every set in R is in the σ-ring generated by the compact subsets of R. Such a space is called *perfect* if a) for any chunk F there is a continuous function on R which vanishes at infinity on R in the topological sense (i. e. for every positive number e, the set where the function is not less than e in absolute value is compact), and is equivalent to the characteristic function of F; b) the measure of any non-void open set in $\mathfrak{R}$ is positive.

Remark 2. 2. It is easy to see that condition b) of the immediately preceding definition is equivalent to the condition that the continuous function described in condition a) be unique.

Examples 2. 1. As examples of localizable spaces we mention a) σ-finite measure spaces, b) a locally compact group relative to Haar measure (see

Corollary 5. 2. 1). To obtain a perfect space, let R have the discrete topology, let $\mathfrak{R}$ consist of all finite subsets of R, and let $r(E)$ be the number of points in E. An example of a space which can be seen without difficulty to be non-localizable is given in [2], p. 131, Note 9.

We conclude this section by showing that either strong or metric equivalence is stronger, actually, than weak equivalence. The following definition is useful in this and other connections.

DEFINITION 2. 8. A Boolean ring is called *countably-decomposable* if any family of mutually disjoint elements of the ring is at most countable. An element of a Boolean ring is called *countably-decomposable* if the principal ideal which it generates is such.

THEOREM 2. 2. *An element of a measure ring is σ-finite if and only if it is countably-decomposable.*

To show that E is countably decomposable if it is a σ-finite element of the measure ring $(\mathfrak{M}, m)$ of the measure space M, assume the contrary. Then $\bar{E} = \bigcup_i \bar{E}_i$, where E_i is a chunk $(i = 1, 2, \cdots)$ and there exists a collection $\{F_\xi; \xi \varepsilon \Xi\}$ of measurable sets in M such that $0 < \bar{F}_\xi \leq \bar{E}$, $\bar{F}_\xi \cap \bar{F}_{\xi'} = 0$ if $\xi \neq \xi'$, and with the index set Ξ more than countable. It follows that $\bar{F}_\xi = \bigcup_i \bar{F}_\xi \cap \bar{E}_i$, for any ξ, and hence that the collection of all $\bar{F}_\xi \cap \bar{E}_i$, as ξ and i vary, is more than countable. It is easy to conclude that for some positive integer n, and some integer i_0, the set $[\xi \mid m(\bar{F}_\xi \cap \bar{E}_{i_0}) > n^{-1}]$ is more than countable. This contradicts the circumstance that the $\bar{F}_\xi \cap \bar{E}_{i_0}$ are mutually disjoint and bounded by the element $\bar{E}_{i_0}$ of finite measure.

Now if $\bar{E}$ is a countably-decomposable element of $\mathfrak{M}$, let $\{\bar{F}_\xi; \xi \varepsilon \Xi\}$ be a maximal family of mutually disjoint elements of $\mathfrak{M}$ of finite measure which are not greater than $\bar{E}$; such a family exists by the Zorn principle, and Ξ is at most countable. The maximality of this family implies that $\bar{E} - \bigcup_\xi \bar{F}_\xi$ is not greater than any element of $\mathfrak{M}$ of finite measure, from which it follows that $\bar{E} = \bigcup_\xi \bar{F}_\xi$.

COROLLARY 2. 2. 1. *The contraction of a strong equivalence to the ideal of σ-finite elements is a weak equivalence. In particular, strongly equivalent spaces are weakly equivalent.*

Since countable decomposability is defined in purely algebraic terms, an algebraic isomorphism maps countably-decomposable elements into countably-decomposable elements. By the theorem just proved, this implies that the contraction of a strong equivalence maps the σ-finite elements onto the σ-finite elements.

THEOREM 2.3. *Let $\mathfrak{F}$ and $\mathfrak{F}'$ be the finite measure rings, and $\mathfrak{S}$ and $\mathfrak{S}'$ the σ-finite measure rings, of the measure spaces M and M' respectively. Any ring-isomorphism of $\mathfrak{F}$ onto $\mathfrak{F}'$ can be uniquely extended to a ring-isomorphism of $\mathfrak{S}$ onto $\mathfrak{S}'$.*

We put η for a ring isomorphism of $\mathfrak{F}$ onto $\mathfrak{F}'$, and define a mapping ψ on $\mathfrak{S}$ to $\mathfrak{S}'$ as follows: if $\bar{E} \in \mathfrak{S}$ and $\bar{E} = \bigcup_i \bar{E}_i$, where $\bar{E}_i$ $(i = 1, 2, \cdots)$ is in $\mathfrak{F}$, then $\psi(\bar{E}) = \bigcup_i \eta(\bar{E}_i)$. Clearly $\psi(\bar{E})$ is defined for every $\bar{E}$ in $\mathfrak{S}$, and we now show that ψ coincides on $\mathfrak{F}$ with η. Let $\bar{E}$ be arbitrary in $\mathfrak{F}$, and suppose that $\bar{E} = \bigcup_i \bar{F}_i$, where $\bar{F}_i$ $(i = 1, 2, \cdots)$ is in $\bar{\mathfrak{F}}$. Then $\bar{E} \geq \bar{F}_i$, whence $\eta(\bar{E}) \geq \eta(\bar{F}_i)$, and $\eta(\bar{E}) \geq \bigcup_i \eta(\bar{F}_i) = \psi(\bar{E})$. Setting $\bar{G}' = \eta(\bar{E}) - \psi(\bar{E})$, then $\bar{G}' \cap \eta(\bar{E}) = \bar{G}'$ and $\bar{G}' \cap \eta(\bar{F}_i) = 0$ $(i = 1, 2, \cdots)$. Now $\bar{G}' = \eta(\bar{G})$, for some $\bar{G}$ in $\mathfrak{F}$, and it results from the equations involving $\bar{G}'$ that $\bar{G} \cap \bar{E} = \bar{G}$ and $\bar{G} \cap \bar{F}_i = 0$ for all i, whence $\bar{G} = \bar{G} \cap \bigcup_i \bar{F}_i = 0$.

Next we show that ψ is single-valued. Let $\bar{E}_i$ and $\bar{F}_i$ $(i = 1, 2, \cdots)$ be in $\mathfrak{F}$, and suppose that $\bigcup_i \bar{E}_i = \bigcup_i \bar{F}_i = \bar{D}$. Let $\bigcup_i \eta(\bar{E}_i) = \bar{A}'$ and $\bigcup_i \eta(\bar{F}_i) = B'$. To show that $\bar{A}' = B'$, it suffices to show that $\bar{A}' \cap \bar{H}' = B' \cap \bar{H}'$ for an arbitrary $\bar{H}'$ in $\mathfrak{F}'$. If $\bar{H}$ is arbitrary in $\mathfrak{F}$, then $\eta(\bar{H})$ is arbitrary in $\mathfrak{F}'$, and hence we need only show that $\eta(\bar{H}) \cap \bigcup_i \eta(\bar{E}_i) = \eta(\bar{H}) \cap \bigcup_i \eta(\bar{F}_i)$. Now

$$\eta(\bar{H}) \cap \bigcup_i \eta(\bar{E}_i) = \bigcup_i \eta(\bar{H}) \cap \eta(\bar{E}_i) = \bigcup_i \eta(\bar{H} \cap \bar{E}_i)$$
$$= \psi(\bigcup_i \bar{H} \cap \bar{E}_i) = \psi(\bar{H} \cap \bar{D}),$$

which equals $\eta(\bar{H} \cap \bar{D})$ since $\bar{H} \cap \bar{D} \in \mathfrak{F}$. As the last expression is independent of the representation of $\bar{D}$ as a countable union of elements of $\mathfrak{F}$, it follows that ψ is single-valued.

To show ψ to be a ring homomorphism, it suffices to show that $\psi(\bar{A} \cap \bar{B}) = \psi(\bar{A}) \cap \psi(\bar{B})$ and $\psi(\bar{A} \cap \bar{B}) = \psi(\bar{A}) \cap \psi(\bar{B})$, for arbitrary $\bar{A}$ and $\bar{B}$ in $\mathfrak{S}$, say $\bar{A} = \bigcup_i \bar{A}_i$ and $\bar{B} = \bigcup_i \bar{B}_i$, where $\bar{A}_i$ and $\bar{B}_i$ $(i = 1, 2, \cdots)$ are in $\mathfrak{F}$. Now

$$\psi(\bar{A} \cap \bar{B}) = \psi(\bigcup_{i,j} \bar{A}_i \cap \bar{B}_j) = \bigcup_{i,j} \eta(\bar{A}_i \cap \bar{B}_j) = \bigcup_{i,j} \eta(\bar{A}_i) \cap \eta(\bar{B}_j)$$
$$= \{\bigcup_i \eta(\bar{A}_i)\} \cap \{\bigcup_j \eta(\bar{B}_j)\} = \psi(\bar{A}) \cap \psi(\bar{B}).$$

In a similar fashion it follows that $\psi(\bar{A} \cap \bar{B}) = \psi(\bar{A}) \cup \psi(\bar{B})$. To see that ψ is onto $\mathfrak{S}'$, let E' be arbitrary in $\mathfrak{S}'$, and let $\bar{E}' = \bigcup_i \bar{E}_i'$, with $\bar{E}_i' \in \mathfrak{F}'$. It is easy to conclude that $\bar{E}' = \psi(\bar{E})$, where $\bar{E} = \bigcup_i \eta^{-1}(\bar{E}_i)$. To complete the proof that ψ is an isomorphism, it remains only to show that ψ is univalent; this is easily seen, for if $\psi(\bar{A}) = \psi(\bar{B})$ with $\bar{A}$ and $\bar{B}$ in $\mathfrak{S}$, then $\psi(\bar{A} \ominus \bar{B}) = 0$, whence $\bar{A} \ominus \bar{B} = 0$, or $\bar{A} = \bar{B}$.

Finally ψ is unique, for let ψ_1 be any ring isomorphism of $\mathfrak{S}$ onto $\mathfrak{S}'$ which extends η. If $\bar{E}$ is arbitrary in $\mathfrak{S}$, say $\bar{E} = \bigcup_i \bar{E}_i$, with $\bar{E}_i \in \mathfrak{J}$ $(i = 1, 2, \cdots)$, then $\bar{E} \geq \bar{E}_i$, $\psi_1(\bar{E}) \geq \psi_1(\bar{E}_i) = \eta(\bar{E}_i)$, whence

$$\psi_1(\bar{E}) \geq \bigcup_i \eta(\bar{E}_i) = \psi(\bar{E}).$$

By the same argument applied to ψ_1^{-1}, ψ^{-1}, and η^{-1} we have $\psi_1^{-1}(\bar{E}') \geq \psi^{-1}(\bar{E}')$, for $\bar{E}' \in \mathfrak{S}'$. Putting $\bar{E}' = \psi(\bar{E})$, it results that $(\psi_1^{-1}\psi)(\bar{E}) \geq \bar{E}$, which implies on applying ψ_1 to both sides of the inequality, that $\psi(\bar{E}) \geq \psi_1(\bar{E})$. Combining the two inequalities shows that $\psi = \psi_1$.

COROLLARY 2. 3. 1. *Metrically equivalent measure spaces are likewise weakly equivalent.*

3. Direct sums of measure spaces. In order to treat measure spaces which are not necessarily σ-finite, it is frequently convenient to represent such spaces in terms of finite spaces. This section is concerned mainly with representations of a measure space as a "direct sum" of other measure spaces, especially of finite ones, where "direct sum" is defined as follows.

DEFINITION 3. 1. Let $M_\xi = (R_\xi, \mathfrak{R}_\xi, r_\xi)$ be a family of measure spaces depending on an index ξ which ranges over a set Ξ, and suppose that R_ξ and $R_{\xi'}$ are disjoint when $\xi \neq \xi'$. The *direct sum* $M = (R, \mathfrak{R}, r)$ of the M_ξ as ξ ranges over Ξ, is the measure space for which $R = \bigcup_\xi R_\xi$; $\mathfrak{R}$ is the collection of all subsets E of R such that $E \cap R_\xi = 0$ except for an at most countable set Ξ_0 of ξ's, and for which $\sum_{\xi \in \Xi_0} r_\xi(E \cap R_\xi)$ is convergent; and r is given by the equation $r(E) = \sum_{\xi \in \Xi_0} r_\xi(E \cap R_\xi)$. The direct sum is denoted as $\Sigma_\xi \oplus M_\xi$.

THEOREM 3. 1. *If the measure space $M_\xi = (R_\xi, \mathfrak{R}_\xi, r_\xi)$ is weakly, or strongly, or metrically, respectively equivalent to the measure space $M_\xi' = (R_\xi', \mathfrak{R}_\xi', r_\xi')$, for all ξ in Ξ, where $R_{\xi_1} \cap R_{\xi_2} = 0 = R_{\xi_1}' \cap R_{\xi_2}'$ for any two distinct elements ξ_1 and ξ_2 of Ξ, then the direct sum of the M_ξ $(\xi \in \Xi)$ is weakly, or strongly, or metrically equivalent to the direct sum of the M_ξ' $(\xi \in \Xi)$.*

Consider first the case of weak equivalence. Let $\mathfrak{F}_\xi$ and $\mathfrak{F}_\xi'$ be the σ-finite measure rings of M_ξ and M_ξ' respectively, and $\mathfrak{F}$ and $\mathfrak{F}'$ the σ-finite measure rings of the direct sums $M = (R, \mathfrak{R}, r)$ and $M' = (R', \mathfrak{R}, r')$ of the M_ξ and the M_ξ' respectively. Let ϕ_ξ be an algebraic isomorphism of $\mathfrak{F}_\xi$ onto $\mathfrak{F}_\xi'$, and let E be an arbitrary σ-finite measurable set in M. Then $\bar{E}$ is an arbitrary element of $\mathfrak{F}$. To show that M and M' are weakly equi-

valent amounts showing that $\mathfrak{F}$ and $\mathfrak{F}'$ are algebraically isomorphic. We define an isomorphism ϕ from $\mathfrak{F}$ onto $\mathfrak{F}'$ as follows. Set $E = \bigcup_{\xi \in \Xi_1} E_\xi$, where Ξ_1 is countable, and $E_\xi = E \cap R_\xi$; then $\bar{E} = \bigcup_{\xi \in \Xi_1} \bar{E}_\xi$ and we define $\phi(\bar{E}) = \bigcup_{\xi \in \Xi_1} \phi_\xi(\bar{E}_\xi)$. Next we show that ϕ is uniquely defined. In the first place, if $E = \bigcup_{\xi \in \Xi_2} E_\xi$, where Ξ_2 is another countable subset of Ξ, then plainly $\Xi_1 \cap \Xi_2 \supset \Xi_0$, where $\Xi_0 = [\xi \mid \bar{E}_\xi \neq 0]$, and it is trivial that $\bigcup_{\xi \in \Xi_1} \phi_\xi(\bar{E}_\xi) = \bigcup_{\xi \in \Xi_0} \phi_\xi(\bar{E}_\xi) = \bigcup_{\xi \in \Xi_2} \phi_\xi(\bar{E}_\xi)$. Furthermore, if F is a measurable σ-finite set such that $\bar{E} = \bar{F}$, then $\bar{E} \cap \bar{R}_\xi = \bar{F} \cap \bar{R}_\xi$, and putting

$$\Xi' = [\xi \mid \text{either } \bar{E} \cap \bar{R} \neq 0 \text{ or } \bar{F} \cap \bar{R}_\xi \neq 0],$$

it results that

$$\phi(\bar{E}) = \bigcup_{\xi \in \Xi'} \phi_\xi(\bar{E} \cap \bar{R}_\xi) = \bigcup_{\xi \in \Xi'} \phi_\xi(\bar{F} \cap \bar{R}_\xi) = \phi(\bar{F}).$$

Thus ϕ is well-defined, and to conclude the proof that it is an isomorphism it remains only to show that it is order-preserving and onto $\mathfrak{F}'$. It is straightforward to verify that ϕ has these further properties.

The proof is the same in the case of metric equivalence except for the additional feature that ϕ_ξ is then taken to be measure-preserving and that ϕ must then be shown to be likewise measure-preserving. It is easily seen that ϕ is actually of this character.

Finally we treat the case of strong equivalence. Let $\mathfrak{m}_\xi$ and $\mathfrak{m}'_\xi$ be the measure rings of M_ξ and M'_ξ, and $\mathfrak{m}$ and $\mathfrak{m}'$ the measure rings of M and M', respectively. Let ψ_ξ be an algebraic isomorphism of $\mathfrak{m}_\xi$ onto $\mathfrak{m}'_\xi$, and let K be an arbitrary measurable set in M. We define an algebraic isomorphism ψ of $\mathfrak{m}$ onto $\mathfrak{m}'$, by setting $\psi(\bar{K}) = \bigcup_\xi \psi_\xi (\bar{K} \cap \bar{R}_\xi)$; we proceed to show that the indicated least upper bound exists. In fact, if L_ξ' is any measurable set in M_ξ', $\bigcup_\xi \bar{L}_\xi'$ exist and equals $\theta(\bigcup_\xi L_\xi')$. For it is easily verified that $\bigcup_\xi L_\xi'$ is measurable and that $\theta(\bigcup_\xi L_\xi') \geq \bar{L}_\xi'$; and if W' is any measurable set in M' such that $\bar{W}' \geq \bar{L}_\xi'$, then $\bar{W}' \cap \bar{R}_\xi' \geq \bar{L}_\xi' \cap \bar{R}_\xi' = \bar{L}_\xi'$. It follows that $\{\bar{W}' - \theta(\bigcup_\xi L_\xi')\} \cap \bar{R}_\xi' = 0$, whence for any subset Ξ_0 of Ξ which is not more than countable,

$$\{\bar{W}' - \theta(\bigcup_\xi L_\xi')\} \cap \bigcup_{\xi \in \Xi_0} \bar{R}_\xi' = 0 = \{\bar{W}' - \theta(\bigcup_\xi L_\xi')\} \cap \theta(\bigcup_{\xi \in \Xi_0} R_\xi').$$

Since every set E' in $\mathfrak{R}'$ is contained in a set of the form $\bigcup_{\xi \in \Xi_0} R_\xi'$, it results that $\{W' - \theta(\bigcup_\xi L_\xi')\} \cap \bar{E}' = 0$ for all such E', which implies that $W' - \bigcup_\xi L_\xi'$ is a null set, i. e. $\bar{W}' \geq \theta(\bigcup_\xi L_\xi')$. It remains only to show that ψ is onto and order-preserving. That ψ has these properties is easily verified.

Direct sums of finite measure spaces occur frequently (e. g. any σ-finite measure space is such) and have the useful property (localizability) stated in the following theorem. Later, Theorem 5. 1, it will be seen that this property almost characterizes direct sums of finite spaces.

THEOREM 3. 2. *A measure space which is strongly equivalent to a direct sum of finite measure spaces is localizable.*

It is plainly sufficient to show that a direct sum of finite measure spaces is localizable. Consider to begin with the case of a single summand.

LEMMA 3. 2. 1. *A finite measure space is localizable.*

Let $\mathcal{B}$ be any subset of the measure ring $(\mathfrak{M}, m)$ of a finite measure space, and let α be the g. l. b. of the measures of all elements of $\mathfrak{M}$ which are upper bound for $\mathcal{B}$,—by the finiteness of the measure space, such elements exist. Let $\{\bar{E}_i; i = 1, 2, \cdots\}$ be a sequence of upper bounds for $\mathcal{B}$ such that $m(\bar{E}_i) \to \alpha$. Putting $\bar{F} = \bigcap_{i \leq n} \bar{E}_i$, it is clear that $\bar{F}_n$ is an upper bound for $\mathcal{B}$, and that $\alpha \leq m(\bar{F}_n) \leq m(\bar{E}_n)$. It follows that $m(\bar{F}_n) \to \alpha$, and hence that $m(\bar{F}) = \alpha$, where $\bar{F} = \bigcap_n \bar{F}_n$, and so $\bar{F}$ is an upper bound for $\mathcal{B}$. Now $\bar{F}$ is the l. u. b. of $\mathcal{B}$, for if $\bar{G}$ is any other upper bound for $\mathcal{B}$, then so is $\bar{F} \cap \bar{G}$, which implies that $m(\bar{F} \cap \bar{G}) = \alpha = m(\bar{F})$, from which it results that $\bar{F} \cap \bar{G} = \bar{F}$ and hence that $\bar{F} \leq \bar{G}$.

The following lemma obviously completes the proof of the theorem.

LEMMA 3. 2. 2. *A direct sum of finite measure spaces is localizable.*

Let M be the direct sum of the measure spaces $M_\xi = (R_\xi, \mathfrak{R}_\xi, r_\xi)$, with ξ ranging over Ξ, and let $\mathcal{B}$ be any collection of elements in the measure ring $\mathfrak{M}$ of M. Let $\mathcal{B}_\xi = [\bar{U} \cap \bar{R}_\xi \mid \bar{U} \in \mathcal{B}]$; then by the preceding lemma $\mathcal{B}_\xi$ has a l. u. b. $\bar{V}_\xi$, where V_ξ can be taken to be a measurable set in M_ξ. Setting $V = \bigcup_\xi V_\xi$, it is easily seen that V is measurable, and as seen in the proof of the preceding theorem, $\bar{V} = \bigcup_\xi \bar{V}_\xi$. Now $\bar{V}$ is a bound for $\mathcal{B}$, for if $\bar{U} \in \mathcal{B}$, then $U = \bigcup_\xi (U \cap R_\xi)$ and again by a fact established in the proof of the preceding theorem, $\bar{U} = \bigcup_\xi (\bar{U} \cap \bar{R}_\xi)$; and as $\bar{V}$ bounds $\mathcal{B}_\xi$, we have in particular, $\bar{V} \geq \bar{U} \cap \bar{R}_\xi$ for all ξ, so that $\bar{V} \geq \bigcup_\xi (\bar{U} \cap \bar{R}_\xi) = \bar{U}$. Moreover, $\bar{V}$ is the l. u. b. of $\mathcal{B}$, for if $\bar{W}$ is any upper bound for $\mathcal{B}$, then it is an upper bound for $\mathcal{B}_\xi$, for each ξ, hence an upper bound for all the $\bar{V}_\xi$ and thus an upper bound for $\bar{V}$.

COROLLARY 3. 2. 1. *A σ-finite measure space is localizable.*

It is clear that a σ-finite measure space is a direct sum of a countable number of finite measure spaces.

The converse of the preceding theorem is included in Theorem 3. 4. The next theorem however is needed in the proof of Theorem 3. 4, and it has some independent interest. It shows that the classification of localizable spaces under strong equivalence is equivalent to their classification under weak equivalence which in turn is reducible (via Theorem 3. 4) to the classification of finite measures spaces under weak equivalence. The latter classification exists in a cogent form in [3a].

THEOREM 3. 3. *Let $\mathfrak{M}$ and $\mathfrak{M}'$ be the measure rings of localizable measure spaces M and M', and let ϕ be an algebraic isomorphism of the σ-finite measure ring of M onto that of M'. Then ϕ can be uniquely extended to an algebraic isomorphism of $\mathfrak{M}$ onto $\mathfrak{M}'$.*

LEMMA 3. 3. 1. *Any element $\bar{K}$ of the measure ring of a measure space is the least upper bound of the elements of finite measure which it bounds.*

It is clear that $\bar{K}$ is an upper bound for the $\bar{K} \cap \bar{E}$, as E ranges over $\mathfrak{R}$, $(R, \mathfrak{R}, r)$ being the measure space in question. To show that it is least, suppose that L is a measurable set in the measure space such that $\bar{L} \geq \bar{K} \cap \bar{E}$, for all $E \in \mathfrak{R}$. Then we have $(\bar{K} - \bar{L}) \cap \bar{E} = (\bar{K} \cap \bar{E}) - (\bar{L} \cap \bar{E}) = (\bar{K} \cap \bar{E}) - (\bar{K} \cap \bar{E}) = 0$. It follows that $K - L$ meets every element of $\mathfrak{R}$ in a null set, i. e. is a null set so $\theta(K - L) = 0$. Now $\theta(K - L) = \bar{K} - \bar{L}$, and hence $\bar{K} \leq \bar{L}$.

To conclude the proof of the theorem, it plainly suffices to establish the following lemma. The lemma has been formulated so as to be applicable to the proof of Theorem 3. 4, and for this reason its statement is slightly more complicated than would be necessary if it required only in connection with the present theorem.

LEMMA 3. 3. 2. *Net $\mathfrak{N}$ and $\mathfrak{N}'$ be Boolean rings with units and let $\mathfrak{D}$ and $\mathfrak{D}'$ be ideals in $\mathfrak{N}$ and $\mathfrak{N}'$, respectively, with the properties that any element of $\mathfrak{N}$ [or $\mathfrak{N}'$] is the least upper bound of the elements of $\mathfrak{D}$ [or $\mathfrak{D}'$] which it bounds. Let ϕ be an algebraic isomorphism of $\mathfrak{D}$ onto $\mathfrak{D}'$ and suppose that, as a lattice, $\mathfrak{N}'$ is complete. Then ϕ can be uniquely extended to an algebraic isomorphism of $\mathfrak{N}$ into $\mathfrak{N}'$, and if $\mathfrak{N}$ is complete, this isomorphism is necessarily onto.*

We define a mapping ψ on $\mathfrak{N}$ to $\mathfrak{N}'$ as follows: for an arbitrary element A of $\mathfrak{N}$, $\psi(A) = \text{l. u. b. } [X \in \mathfrak{D}, X \leq A]\phi(X)$. It is clear that ψ is order-preserving and agrees on $\mathfrak{D}$ with ϕ. We proceed to show that it is univalent (i. e. one-to-one into) and a homomorphism.

To prove that ψ is univalent, it is sufficient to show that if $\psi(A) \geqq \psi(B)$,

then $A \geq B$. Now assuming $\psi(A) \geq \psi(B)$, if $A \geq B$ is false, there is a non-zero element C of $\mathfrak{D}$ such that $B - A \geq C$, since otherwise $(B - A) \cap X = 0$ for all $X \in \mathfrak{D}$ and from the hypothesis, $B - A = \text{l. u. b.}_{X \in \mathfrak{Q}}(B - A) \cap X = 0$, or $A \geq B$. Thus $B \geq C$, $\psi(B) \geq \psi(C)$, and hence

$$\psi(A) \cap \psi(C) \geq \psi(B) \cap \psi(C) = \psi(C).$$

On the other hand,

$$\begin{aligned}\psi(A) \cap \psi(C) &= \psi(C) \cap \text{l. u. b. } [X \in \mathfrak{D} \mid X \leq A]\phi(X) \\ &= \text{l. u. b. } [X \in \mathfrak{D}, X \leq A]\{\phi(C) \cap \phi(X)\} \\ &= \text{l. u. b. } [X \in \mathfrak{D}, X \leq A]\phi(C \cap X).\end{aligned}$$

The last expression vanishes because when $X \leq A$, $C \cap X \leq C \cap A = 0$. Therefore $\psi(C) = 0$, *i. e.* $\phi(C) = 0$, and since ϕ is an isomorphism, there results the contradiction $C = 0$.

To show that ψ is a homomorphism, let A and B be arbitrary elements of $\mathfrak{N}$, and let D be an arbitrary element of $\mathfrak{D}$. Then

$$\begin{aligned}\psi(D) \cap \psi(A) &= \psi(D) \cap \text{l. u. b. } [X \in \mathfrak{D}, X \leq A]\phi(X) \\ &= \text{l. u. b. } [X \in \mathfrak{D}, X \leq A]\phi(D \cap X) = \text{l. u. b. } [Y \in \mathfrak{D}, Y \leq D \cap A]\phi(Y) \\ &= \psi(D \cap A).\end{aligned}$$

Now

$$\begin{aligned}&\text{l. u. b. } [D \in \mathfrak{D}, D \leq B](\psi(D) \cap \psi(A)) \\ &\quad = \psi(A) \cap \text{l. u. b. } [D \in \mathfrak{D}, D \leq B]\psi(D) = \psi(A) \cap \psi(B).\end{aligned}$$

On the other hand,

$$\begin{aligned}&\text{l. u. b. } [D \in \mathfrak{D}, D \leq B]\psi(D \cap A) \\ &\quad = \text{l. u. b. } [X \in \mathfrak{D}, X \leq A \cap B]\psi(X) = \psi(A \cap B).\end{aligned}$$

Thus $\psi(A \cap B) = \psi(A) \cap \psi(B)$. A parallel procedure shows that $\psi(A \cup B) = \psi(A) \cup \psi(B)$.

Thus ψ is an algebraic isomorphism of $\mathfrak{N}$ into $\mathfrak{N}'$ and extends ϕ. If ψ_1 is any such isomorphism, and if A is an arbitrary element of $\mathfrak{N}$, the equation $A = \text{l. u. b. } [X \in \mathfrak{D}, X \leq A]X$ implies that

$$\psi_1(A) \geq \text{l. u. b. } [X \in \mathfrak{D}, X \leq A]\phi(X) = \psi(A).$$

In particular, if I and I' are the units of $\mathfrak{N}$ and $\mathfrak{N}'$ respectively, $\psi_1(I) \geq \psi(I) = I'$, so that $\psi_1(I) = I'$. It follows that $\psi_1(I - A) = I' - \psi_1(A)$, and employing this equation in conjunction with the inequality $\psi_1(I - A)$

$\geq \psi(I - A)$ shows that $\psi(A) \geq \psi_1(A)$. As the opposite semi-inequality has already been obtained, it follows that $\psi = \psi_1$, i. e. ψ has the stated uniqueness property.

Now suppose that $\mathfrak{N}$ is complete. To show that ψ is then onto, let A' be an arbitrary element of $\mathfrak{N}'$; we shall show that there exists an element A of $\mathfrak{N}$ such that $\psi(A) = A'$. Set $A = \text{l. u. b.}\,[X \in \mathfrak{D}, \phi(X) \leq A']X$. Then $A \geq X$ for $X \in \mathfrak{D}$ and $\phi(X) \leq A'$, whence $\psi(A) \geq \phi(X)$ for all such X. Since ϕ is onto, A' is the l. u. b. of the $\phi(X)$, for $X \in \mathfrak{D}$ and $\phi(X) \leq A'$, and hence $\psi(A) \geq A'$. On the other hand, if D' is an element of $\mathfrak{D}'$ such that $D' \leq \psi(A) - A'$, then $D' \leq \psi(A)$, i. e. $D' \leq \text{l. u. b.}\,[X \in \mathfrak{D}, X \leq A]\phi(X)$, and intersecting this inequality with D' yields the equation

$$D' = \text{l. u. b.}\,[X \in \mathfrak{D}, X \leq A]\{\phi(X) \cap \phi(D)\},$$

where D is the element of $\mathfrak{D}$ for which $\phi(D) = D'$. Now $\phi(X) \cap \phi(D) = \phi(X \cap D)$, and as X ranges over the elements of $\mathfrak{D}$ which are bounded by A, $X \cap D$ ranges over the elements of $\mathfrak{D}$ which are bounded by $A \cap D$. It results that $D' = \text{l. u. b.}\,[Y \in \mathfrak{D}, Y \leq A \cap D]\phi(Y)$. Intersecting the equation, $A = \text{l. u. b.}\,[\phi(X) \leq A', X \in \mathfrak{D}]$, with D yields the equation $A \cap D = \text{l. u. b.}\,[\phi(X) \leq A', X \in \mathfrak{D}]X \cap D$. If $X \in \mathfrak{D}$ and $\phi(X) \leq A'$, then since $D' \leq \psi(A) - A'$ we have $D' \cap A' = 0$, and hence $D' \cap \phi(X) = 0 = \phi(X \cap D)$. This implies $X \cap D = 0$, and hence it follows from the equation for $A \cap D$ that $A \cap D = 0$. The equation for D' then shows that $D' = 0$. Thus $\{\psi(A) - A'\} \cap D' = 0$ for all $D' \in \mathfrak{D}'$. Since any element of $\mathfrak{N}'$ is the l. u. b. of the elements of $\mathfrak{D}'$ which it bounds,

$$\psi(A) - A' = \text{l. u. b.}_{D' \in \mathfrak{D}'}\{\psi(A) - A'\} \cap D' = 0.$$

In other terms, $A' \geq \psi(A)$, and since the semi-inequality in the opposite direction has been previously obtained, we can conclude that $A' = \psi(A)$.

The following theorem is useful for reducing questions concerning arbitrary measure spaces to corresponding questions for finite measure spaces. In combination with Theorem 3. 2 it shows that localizable spaces are precisely those which are, in some ways, most readily treated in terms of finite spaces, and thereby helps to justify our designation for these spaces (see also condition *e.* of Theorem 5. 1).

THEOREM 3. 4. *For any measure space M there is a measure-preserving algebraic isomorphism of its measure ring $\mathfrak{M}$ into the measure ring $\mathfrak{M}'$ of a direct sum of finite measure spaces, whose contraction to the σ-finite subring of $\mathfrak{M}$ is onto the same subring of $\mathfrak{M}'$. For such an isomorphism to be onto it is necessary and sufficient that M be localizable.*

Let $M = (R, \mathcal{R}, r)$ and let $\{R_\xi; \xi \in \Xi\}$ be a family of chunks in M of positive measure, which is maximal with respect to the property that the intersection of any two sets in the family is of measure zero,—it is clear from Zorn's principle that such a family exists. Let $\mathcal{R}_\xi$ be the subring of $\mathcal{R}$ consisting of the elements of $\mathcal{R}$ which are contained in R_ξ, and let $M_\xi = (R_\xi, \mathcal{R}_\xi, r)$. Let $M_\xi' = (R_\xi', \mathcal{R}_\xi', r_\xi')$ be a space isomorphic to M_ξ whose set R_ξ' is chosen in such a way that $R_{\xi_1}' \cap R_{\xi_2}' = 0$ for $\xi_1 \neq \xi_2$—e. g., R_ξ' may be taken to be the set of all pairs (x, ξ) with $x \in R_\xi$, and with $\mathcal{R}_\xi'$ and r_ξ' determined by the condition that the mapping η_ξ, $x \to (x, \xi)$, on R_ξ to R_ξ', be an isomorphism on M_ξ to M_ξ'. Now let $M' = (R', \mathcal{R}', r')$ be the direct sum of the M_ξ', $\xi \in \Xi$. It is easily seen that if E is in the family $\mathcal{E}$ of all chunks in M, then the set $\Xi_0 = [\xi \mid r(E \cap R_\xi) \neq 0]$ is at most countable. Since $\sum_{\xi \in \Xi_0} r(E \cap R_\xi) \leq r(E)$, the sum $\sum_{\xi \in \Xi_0} r_\xi'(\eta_\xi(E \cap R_\xi))$ is convergent, and hence $\bigcup_{\xi \in \Xi_0} \eta_\xi(E \cap R_\xi)$, $= E'$ say, is in $\mathcal{R}'$. We define a mapping ϕ_0 on $\mathcal{E}$ to $\mathcal{R}'$ as follows: $\phi_0(E) = E'$, and observe that if E and F are equivalent elements of $\mathcal{E}$, then $\phi_0(E)$ and $\phi_0(F)$ are equivalent elements of $\mathcal{R}'$. It results that the mapping ϕ, on $\mathfrak{D} = \theta(\mathcal{E})$ to $\mathfrak{D}' = \theta(\mathcal{R}')$, defined by the equation: $\phi(\bar{E}) = \theta(\phi_0(E))$, $E \in \mathcal{E}$, is single-valued. We show next that ϕ is a measure-preserving algebraic isomorphism of $\mathfrak{D}$ onto $\mathfrak{D}'$.

That ϕ is measure-preserving means that if E is as in the preceding paragraph, then $r(E) = r'(E')$. Now

$$r'(E') = \sum_{\xi \in \Xi_0} r_\xi'(\eta_\xi(E \cap R_\xi)) = \sum_{\xi \in \Xi_0} r(E \cap R_\xi).$$

Putting $G = E - \bigcup_{\xi \in \Xi_0} (E \cap R_\xi)$, it follows that G is a chunk and has measure $r(E) - r'(E')$. It is easy to see that $G \cap R_\xi$ is a null set, for all ξ. The maximality of the collection $\{R_\xi; \xi \in \Xi_0\}$ hence implies that G is a null set, i. e. $r(E) - r'(E') = 0$.

To show that ϕ is a homomorphism, let E_1 and E_2 be arbitrary in $\mathcal{E}$, and let $\Xi_3 = \Xi_1 \cup \Xi_2$, where $\Xi_i = [\xi \mid r(E_i \cap R_\xi) \neq 0]$ $(i = 1, 2)$. Observe that if E is arbitrary in $\mathcal{E}$, as above, and if Ξ_4 is any set such that $\Xi_4 \supset \Xi_3 \cup \Xi_0$ then $\phi(\bar{E}) = \bigcup_{\xi \in \Xi_4} \theta(\eta_\xi(E \cap R_\xi))$, for $E \cap R_\xi$ is a null set when $\xi \in \Xi_4 - \Xi_0$, so that for these values of ξ, $\theta(\eta_\xi(E \cap R_\xi)) = 0$, and hence $\bigcup_{\xi \in \Xi_4} \theta(\eta_\xi(E \cap R_\xi)) = \bigcup_{\xi \in \Xi_0} \theta(\eta_\xi(E \cap R_\xi)) = \theta(\bigcup_{\xi \in \Xi_0} \eta_\xi(E \cap R_\xi)) = \bar{E}'$. It follows that $\phi(\bar{E}_1) \cup \phi(\bar{E}_2) = \{\bigcup_{\xi \in \Xi_3} \theta(\eta_\xi(E_1 \cap R_\xi))\} \cup \bigcup_{\xi \in \Xi_3} \theta(\eta_\xi(E_2 \cap R_\xi))$. Now M' is a direct sum of finite measure spaces, and hence localizable, by Theorem 3.2.

i. e. its measure ring $\mathfrak{M}'$ is complete. Consequently the right side of the last equation can be expressed in the form

$$\bigcup_{\xi \in \Xi_3} \{\theta(\eta_\xi(E_1 \frown R_\xi)) \smile \theta(\eta_\xi(E_2 \frown R_\xi))\}$$
$$= \bigcup_{\xi \in \Xi_3} \theta(\eta_\xi(E_1 \frown R_\xi) \smile \eta_\xi(E_2 \frown R_\xi))$$
$$= \bigcup_{\xi \in \Xi_3} \theta(\eta_\xi((E_1 \smile E_2) \frown R_\xi) = \phi(\theta(E_1 \smile E_2)) = \phi(\bar{E}_1 \smile \bar{E}_2).$$

In a similar fashion it can be shown that $\phi(\bar{E}_1) \frown \phi(\bar{E}_2) = \phi(\bar{E}_1 \frown \bar{E}_2)$.

It follows that ϕ is an isomorphism, for if $\phi(\bar{E}_1) = \phi(\bar{E}_2)$, where E_1 and E_2 are as above, then $\phi(\bar{E}_1 \ominus \bar{E}_2) = 0$, so that $\bar{E}_1 \ominus \bar{E}_2 = 0$, or $\bar{E}_1 = \bar{E}_2$. Moreover, ϕ is onto $\mathfrak{D}'$, for if F' is in $\mathfrak{R}'$, $F' = \bigcup_{\xi \in \Xi_0} (F' \frown R'_\xi)$ for some set Ξ_0 which is at most countable, and setting $F = \bigcup_{\xi \in \Xi_0} \eta_\xi^{-1}(F' \frown R_\xi)$, F is clearly measurable, $r(F) = \sum_{\xi \in \Xi_0} r(\eta_\xi^{-1}(F' \frown R_\xi)) = \sum_{\xi \in \Xi_0} r'(F' \frown R'_\xi) = r(F')$ so that F is a chunk, and it is clear that $\phi(\bar{F}) = \bar{F}'$.

By Lemma 3. 3. 2 (which is applicable because of Lemma 3. 3. 1), ϕ can be uniquely extended to an algebraic isomorphism ψ of $\mathfrak{M}$ into $\mathfrak{M}'$. Now ϕ is measure-preserving, and ψ agrees with ϕ except on elements of $\mathfrak{M} - \mathfrak{D}$, all of which elements have infinite measure and map into elements of $\mathfrak{M}' - \mathfrak{D}'$, which likewise have infinite measure. Therefore ψ is measure-preserving. It is clear that ψ maps σ-finite elements of $\mathfrak{M}$ into σ-finite elements of $\mathfrak{M}'$ Conversely, if $\bar{E}'$ is any σ-finite element of $\mathfrak{M}'$, then we have $\bar{E}' = \bigcup_i \bar{E}'_i$, where $\{E_i; i = 1, 2, \cdots\}$ is a sequence of chunks. Now any chunk H in M differs from a countable union of elements in $\mathfrak{R}$ by a set of measure zero, for if $\{H_n; n = 1, 2, \cdots\}$ is a sequence of elements of $\mathfrak{R}$ such that $r(H_n)$ converges to the least upper bound of the measure of subsets of H which are in $\mathfrak{R}$, then it is easily seen that $H - \bigcup_n H_n$ intersects each element of $\mathfrak{R}$ in a null set and so has measure zero. It follows that E' differs from the union of a sequence $\{F_i'; i = 1, 2, \cdots\}$ of elements of $\mathfrak{R}'$ by a null set. Hence $\bar{E}' = \bigcup_i \bar{F}_i'$ and putting $\bar{E} = \bigcup_i \phi^{-1}(\bar{F}_i)$, it results that $\bar{E}$ is a σ-finite element of $\mathfrak{M}$ such that $\psi(\bar{E}) = \bar{E}'$.

Finally, if M is localizable, then $\mathfrak{M}$ is complete,, and by Lemma 3. 3. 2 ψ is necessarily onto. On the other hand, if ψ is onto, then M' is strongly equivalent to a direct sum of finite measure spaces, and hence by Theorem 3. 2 is localizable.

4. Equivalences of spaces. In the present section we study the relations between various types of equivalence of measure spaces, and corresponding types of equivalence of function spaces or operator algebras associated with

the measure spaces. Before stating our first theorem along these lines we make a number of definitions and introduce some additional notation.

DEFINITIONS AND NOTATIONS 4.1. *Multiplication algebras.* The *multiplication algebra* of a measure space $M = (R, \mathfrak{R}, r)$ is the algebra of all operators T_k on $B_2(M)$ of the form $(T_k f)(x) = k(x)f(x)$, where f is a general element of $L_2(M)$ and k is an arbitrary bounded measurable function on M. Here, and elsewhere when convenient, we simplify notation by denoting a residue class in $B_\alpha(M)$ by any function in that residue class. T_k is called the *operation of multiplication by* k. Two algebras of bounded linear operators $\mathcal{A}$ and $\mathcal{A}'$ on Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ are called *unitarily equivalent* if there exists a unitary transformation U from $\mathcal{H}$ onto $\mathcal{H}'$ such that $U\mathcal{A} = \mathcal{A}'U$, and the mapping $W \to UWU^{-1}$ is then called the *algebraic isomorphism of* $\mathcal{A}$ *induced by* U. If K is any set in M, χ_K denotes its characteristic function, and if K is measurable, $\chi_{\bar{K}}$ denotes the residue class in $B_\infty(M)$ of which χ_K is an element.

Remark 4.1. It is obvious that $\chi_{\bar{K}}$ actually depends only on $\bar{K}$, so that our notation is justified. We mention also that it is easy to verify that the mapping $k \to T_k$ on $B_\infty(M)$ to the multiplication algebra of M is an isomorphism.

The following theorem has interesting implications for multiplicity theory, by virtue of the fact that the multiplication algebra of a localizable measure space is essentially the most general maximal abelian SA (self-adjoint) algebra of bounded linear operators on a Hilbert space (cf. [5]). The theorem also includes a kind of analog, for measure spaces, of the Stone-Gelfand-Kolmogoroff theorem associating a homomorphism between compact Hausdorff spaces with an algebraic isomorphism between function spaces over them.

THEOREM 4.1. *Two measure spaces are strongly equivalent if and only if their algebras of bounded measurable functions are algebraically isomorphic (complex conjugation of functions being taken to be algebraic operation), which in turn is true if and only if their multiplcation algebras are unitarily equivalent.*

More specifically, let M *and* M' *be measure spaces with measure rings* $\mathfrak{M}$ *and* $\mathfrak{M}'$, *algebras of bounded measurable functions* $\mathfrak{B}$ *and* $\mathfrak{B}'$, *multiplication algebras* $\mathcal{A}$ *and* $\mathcal{A}'$. *Then if* ϕ *is a mapping on* $\mathfrak{M}$ *to* $\mathfrak{M}'$ *which is a strong equivalence between* M *and* M', *there is an algebraic isomorphism* $\tilde{\phi}$ *from* $\mathfrak{B}$ *onto* $\mathfrak{B}'$ *such that* $\tilde{\phi}(\chi_K) = \chi_{\phi(K)}$, *for all* $\bar{K}$ *in* $\mathfrak{M}$; *and conversely every algebraic isomorphism* $\tilde{\phi}$ *from* $\mathfrak{B}$ *onto* $\mathfrak{B}'$ *arises in this fashion.*

Furthermore, if $\tilde{\phi}$ is an algebraic isomorphism from $\mathfrak{B}$ onto $\mathfrak{B}'$, then $\mathcal{A}$ and $\mathcal{A}'$ are unitarily equivalent via a unitary transformation for which the induced algebraic isomorphism of $\mathcal{A}$ takes multiplication by k into multiplication by $\tilde{\phi}(k)$, for $k \in \mathfrak{B}$.

It is appropriate to divide the proof into two parts.

PART A: *Strong equivalence and isomorphism of algebras of bounded measurable functions.* Let ϕ be an algebraic isomorphism on $\mathfrak{M}$ to $\mathfrak{M}'$, and let $\mathfrak{L}$ [or $\mathfrak{L}'$] be the linear spaces of all elements of $\mathfrak{B}$ [or $\mathfrak{B}'$, respectively] which are finite linear combinations of characteristic functions of measurable sets. We define a mapping $\tilde{\phi}_0$ on $\mathfrak{L}$ to $\mathfrak{L}'$ as follows: for any mutually disjoint measurable sets E_i $(i = 1, 2, \cdots, n;\ n$ finite) in M, $\tilde{\phi}_0(\Sigma_i \alpha_i \chi_{E_i}) = \Sigma_i \alpha_i \chi_{\phi(E_i)}$. To see that ϕ_0 is single-valued, i. e. does not depend on the manner in which the function $f = \Sigma_i \alpha_i \chi_{E_i}$ is represented, let $\gamma_1, \cdots, \gamma_p$ be the distinct nonzero values which f assumes on the sets (necessarily measurable) $H_0, H_1, \cdots, H_p$ of positive measure. Setting $\Lambda_1 = [i \mid \alpha_i = 0]$, $\Lambda_2 = [i \mid \alpha_i \neq 0$ and $\bar{E}_i = 0]$, $\Lambda_3 = [i \mid \alpha_i \neq 0$ and $\bar{E}_i \neq 0]$, it is clear that

$$\Sigma_i \alpha_i \chi_{\phi(\bar{E}_i)} = \Sigma_{i \in \Lambda_1} \alpha_i \chi_{\phi(\bar{E}_i)} + \Sigma_{i \in \Lambda_2} \alpha_i \chi_{\phi(\bar{E}_i)} + \Sigma_{i \in \Lambda_3} \alpha_i \chi_{\phi(\bar{E}_i)} = \Sigma_{i \in \Lambda_3} \alpha_i \chi_{\phi(\bar{E}_i)}.$$

Now putting $\Omega_j = [i \mid i \in \Lambda_3, \alpha_i = \gamma_j]$ it is plain that the Ω_j are mutually disjoint and cover Λ_3 so that

$$\Sigma_{i \in \Lambda_3} \alpha_i \chi_{\phi(E_i)} = \Sigma_j \Sigma_{i \in \Omega_j} \alpha_i \chi_{\phi(E_i)} = \Sigma_j \gamma_j \Sigma_{i \in \Omega_j} \chi_{\phi(E_i)} = \Sigma_j \gamma_j \chi_{\bigcup_{i \in \Omega_j} \phi(E_i)}$$
$$= \Sigma_j \gamma_j \chi_{\phi(\bigcup_{i \in \Omega_j} \bar{E}_i)} = \Sigma_j \gamma_j \chi_{\phi(\bar{H}_j)},$$

noting that $\bigcup_{i \in \Omega_j} \bar{E}_i = \bar{H}_j$. Since the last expression for $\tilde{\phi}_0(f)$ is independent of the manner in which f is represented, it follows that $\tilde{\phi}_0$ is single-valued.

It is easy to see that $\mathfrak{L}$ and $\mathfrak{L}'$ are subalgebras of $\mathfrak{B}$ and $\mathfrak{B}'$, and we show next that $\tilde{\phi}_0$ is a homomorphism. Let $f = \Sigma_i \alpha_i \chi_{\bar{E}_i}$ and $g = \Sigma_j \beta_j \chi_{\bar{F}_j}$ be arbitrary elements of $\mathfrak{L}$. Putting $\bar{G}_{ij} = \bar{E}_i \cap \bar{F}_j$, the $\bar{G}_{ij}$ are mutually disjoint, and $\chi_{\bar{E}_i} = \chi_{\bigcup_j \bar{G}_{ij}} = \Sigma_j \chi_{\bar{G}_{ij}}$, $\chi_{\bar{F}_j} = \chi_{\bigcup_i \bar{G}_{ij}} = \Sigma_i \chi_{\bar{G}_{ij}}$. It follows that $\chi_{\phi(E_i)} = \chi_{\phi(\bigcup_j \bar{G}_{ij})} = \chi_{\bigcup_j \phi(\bar{G}_{ij})} = \Sigma_j \chi_{\phi(\bar{G}_{ij})}$ and similarly $\chi_{\phi(\bar{F}_j)} = \Sigma_i \chi_{\phi(\bar{G}_{ij})}$. Therefore $f + g = \Sigma_i \alpha_i \Sigma_j \chi_{\bar{G}_{ij}} + \Sigma_j \beta_j \Sigma_i \chi_{\bar{G}_{ij}} = \Sigma_{i,j} (\alpha_i + \beta_j) \chi_{\bar{G}_{ij}}$, and

$$\tilde{\phi}_0(f + g) = \Sigma_{i,j} (\alpha_i + \beta_j) \chi_{\phi(\bar{G}_{ij})} = \Sigma_i \alpha_i \Sigma_j \chi_{\phi(\bar{G}_{ij})} + \Sigma_j \beta_j \Sigma_i \chi_{\phi(G_{ij})}$$
$$= \Sigma_i \alpha_i \chi_{\phi(\bar{E}_i)} + \Sigma_j \beta_j \chi_{\phi(\bar{F}_j)} = \tilde{\phi}_0(f) + \tilde{\phi}_0(g).$$

It is easy to see also that $\tilde{\phi}_0(fg) = \tilde{\phi}_0(f) \tilde{\phi}(g)$ and that $\tilde{\phi}_0(\alpha f) = \alpha \tilde{\phi}_0(f)$ for any complex number α.

Now if f is as above, $\| f \|_\infty = \text{l. u. b.}_j \mid \gamma_j \mid = \| \tilde{\phi}_0(f) \|$, so that $\tilde{\phi}_0$ is an

isometry. Moreover, $\tilde{\phi}_0$ is onto $\mathfrak{L}'$, for if $f' = \Sigma_i \alpha_i \chi_{\bar{E}'_i}$ is an arbitrary element of $\mathfrak{L}'$, putting $\bar{E}_i$ for the element of $\mathfrak{M}$ such that $\phi(\bar{E}_i) = \bar{E}_i'$ and $f = \Sigma \alpha_i \chi_{\bar{E}_i}$ it is obvious that $\tilde{\phi}_0(f) = f'$. It follows that $\tilde{\phi}_0$ can be extended to an isometry ϕ of the closure of $\mathfrak{L}$ onto the closure of $\mathfrak{L}'$, and it is easily seen that these closures are algebras and that ϕ is a homomorphism. To show that ϕ has the properties alleged in the theorem, it is only necessary now to show that $\mathfrak{L}$ [and $\mathfrak{L}'$] are dense in $\mathfrak{B}$ [and $\mathfrak{B}'$].

It suffices to show that the real elements of $\mathfrak{L}$ are dense in the set of all real elements of $\mathfrak{B}$, for if h is an arbitrary element of $\mathfrak{B}$, $h = h_1 + ih_2$, where h_1 and h_2 are real, and if the h_i can be approximated by real elements of $\mathfrak{L}$, then h can be approximated by complex elements; and since M is an arbitrary measure space, the density of $\mathfrak{L}'$ in $\mathfrak{B}'$ follows from the density of $\mathfrak{L}$ in $\mathfrak{B}$. Now let k be an arbitrary real element in $\mathfrak{B}$, and let ϵ be an arbitrary positive number. Set $E_n = [x \mid n\epsilon < k(x) \leq (n+1)\epsilon]$ $(n = 0, \pm 1, \cdots)$. Then E_n is measurable, and is of measure zero for $|n|$ sufficiently large, say for $|n| > |m|$. Putting $k_\epsilon = \sum_{n=-m}^{m} n\epsilon \chi_{E_n}$ it is easy to conclude that $k_\epsilon \in \mathfrak{L}$, and that $\| k - k_\epsilon \|_\infty \leq \epsilon$.

Remark 4.2. The mapping $\tilde{\phi}$ of Theorem 4.1 is necessarily unique. To see this, observe that any algebraic isomorphism of $\mathfrak{B}$ onto $\mathfrak{B}'$ preserves norms, for the norm of an element k of $\mathfrak{B}$ can be defined purely algebraically, as the g. l. b. of the non-negative numbers α for which there exists an element h of $\mathfrak{B}$ such that $\alpha^2 - k\bar{k} = h\bar{h}$. Thus any such isomorphism is necessarily continuous, and since it must evidently coincide on $\mathfrak{L}$ with $\tilde{\phi}_0$, it is identical with $\tilde{\phi}$.

Completion of proof for Part A. Conversely now, if $\tilde{\phi}$ is an algebraic isomorphism of $\mathfrak{B}$ onto $\mathfrak{B}'$, and if E is any measurable set in M, then from the equation $(\chi_{\bar{E}})^2 = \chi_{\bar{E}}$ it follows that $\tilde{\phi}((\chi_{\bar{E}})^2) = \tilde{\phi}(\chi_{\bar{E}})$, which implies that $\tilde{\phi}(\chi_{\bar{E}}) = \chi_{\bar{E}'}$ for some $\bar{E}' \in \mathfrak{M}'$. We define ϕ as the mapping on $\mathfrak{M}$ to $\mathfrak{M}'$ given by the equation: $\phi(\bar{E}) = \bar{E}'$. Then ϕ is univalent, for if $\phi(\bar{E}) = \phi(\bar{F})$, $\bar{F}$ being in $\mathfrak{M}$, it results that $\chi_{\phi(\bar{E})} = \chi_{\phi(\bar{F})}$, $\tilde{\phi}(\chi_{\bar{E}}) = \tilde{\phi}(\chi_{\bar{F}})$, $\chi_{\bar{E}} = \chi_{\bar{F}}$ and so $\bar{E} = \bar{F}$. Also, ϕ is onto, for if $\bar{F}' \in \mathfrak{M}'$, then since $\tilde{\phi}$ is onto there exists $f \in \mathfrak{B}$ such that $\tilde{\phi}(f) = \chi_{\bar{F}'}$; applying $\tilde{\phi}^{-1}$ to the equation $(\chi_{\bar{F}'})^2 = \chi_{\bar{F}'}$ yields the equation $f^2 = f$, from which it follows that $f = \chi_{\bar{F}}$ for some $\bar{F} \in \mathfrak{M}$, and that $\bar{F}' = \phi(\bar{F})$. Finally, ϕ is order-preserving (and hence an algebraic isomorphism) for if $\bar{F} \in \mathfrak{M}$ and $\bar{F} \leq \bar{E}$, then $\chi_{\bar{F}}\chi_{\bar{E}} = \chi_{\bar{F}}$ and applying $\tilde{\phi}$ to this equation leads to the equation $\chi_{\phi(\bar{F})}\chi_{\phi(\bar{E})} = \chi_{\phi(\bar{F})}$ which implies that $\phi(\bar{F}) \leq \phi(\bar{E})$.

PART B: *Isomorphism of algebras of bounded measurable functions and unitary equivalence of multiplication algebras.*

In this part, $\tilde{\phi}$ is a given algebraic isomorphism of $\mathfrak{B}$ onto $\mathfrak{B}'$, and it must be shown that the algebraic isomorphism Φ of $\mathfrak{A}$ onto $\mathfrak{A}'$, where Φ takes multiplication by k into multiplication by $\tilde{\phi}(k)$, is induced by an appropriate unitary transformation of $L_2(M)$ onto $L_2(M')$. We shall exhibit this unitary transformation as the product of two such transformations. The first of these is from $L_2(M)$ onto $L_2(M'')$, where M'' has the following properties: it is metrically equivalent to M, and on the other hand $M'' = (R', \mathfrak{S}', s')$, where $M' = (R', \mathfrak{R}', r')$, M' and M'' have the same measurable sets, and where s' and r' are absolutely continuous with respect to each other. The second transformation is then from $L_2(M'')$ onto $L_2(M')$.

Let ϕ be the isomorphism of $\mathfrak{M}$ onto $\mathfrak{M}'$ which induces $\tilde{\phi}$ as in Part A. For any measurable set E' in M' we define a measure s' by the equation $s'(E') = r(\phi^{-1}(\bar{E}'))$. It is clear that s' is countably additive on the collection of measurable sets in M', and that if $\mathfrak{S}'$ is the collection of measurable sets E' such that $s'(E')$ is finite, then $M'' = (R', \mathfrak{S}', s')$ is a measure space.

LEMMA 4.1.1. *A set is measurable, or measurable and σ-finite, or a null set in M', if and only if it is the same in M''.*

We observe to begin with any σ-finite measurable set H' in M' is equivalent in M' to a σ-finite measurable set in M''. For $\bar{H}'$ is then σ-finite in $\mathfrak{M}'$, so that $\bar{H}' = \phi(\bar{H})$, where $\bar{H}$ is a σ-finite element of $\mathfrak{M}$, say $\bar{H} = \bigcup_i \bar{H}_i$, where $\bar{H}_i$ has finite measure. It then follows that $\bar{H}' = \bigcup_i \bar{H}_i'$, where $\bar{H}_i'$ is a measurable set in M' such that $\bar{H}_i' = \phi(\bar{H}_i)$, or that H' is equivalent in M' to the union of the H_i'. As each H_i' is in $\mathfrak{S}'$, this union is a σ-finite measurable set in M''.

It is obvious that if E' is a measurable set in M', then it is a measurable set in M''. Now if E' is a measurable set in M'', and if H' is in $\mathfrak{R}'$, then H' is σ-finite in M', and hence by the preceding paragraph, $H' = \bigcup_i H_i' \cup N'$, where $H_i' \in \mathfrak{S}'$ $(i = 1, 2, \cdots)$ and N' is equivalent to zero in M'. Clearly $E' \cap H' = \bigcup_i (E' \cap H_i') \cup (E' \cap N')$, and since $E' \cap H_i' \in \mathfrak{S}'$ and so is measurable in M', it results that $E' \cap H'$ is measurable in M', and hence that E' is measurable in M'.

Now if G' is a σ-finite measurable set in M'', we have $G' = \bigcup_i G_i' \cup N''$, where $G_i' \in \mathfrak{S}'$ and N'' is equivalent to zero in M''. Now $\bar{G}_i'$ (regarding G_i' here as a measurable set in M') is σ-finite, for $s'(\bar{G}_i') = r(\phi^{-1}(\bar{G}_i'))$ is finite, showing that $\phi^{-1}(\bar{G}_i')$ is σ-finite, and hence that $\bar{G}_i'$ is such. Hence G_i' is

equivalent in M' to a σ-finite set in M', whence $G' = F' \cup N''$, where F' is σ-finite in M', and N'' is equivalent to zero in M''. It follows that to conclude the proof of the lemma, it suffices to show that a set is equivalent to zero in M' if and only if it is such in M''.

If E' is equivalent to zero in M', but not in M'', then $E' \supset W'$ where $W' \varepsilon \mathcal{S}'$ and $s'(W') \neq 0$. Now $s'(W') = r'(\phi^{-1}(\bar{W}'))$ (here $\bar{W}' = \theta_{M'}(W')$, i. e. $\bar{W}'$ is the residue class containing W' relative to M') which shows that $\phi^{-1}(\bar{W}') \neq 0$ and hence $\bar{W}' \neq 0$. This contradicts the relations $\bar{W} \leq \bar{E}'$ and $\bar{E}' = 0$. If on the other hand E'' is equivalent to zero in M'', but not in M', then $E'' \supset W'$, where $W' \varepsilon \mathcal{R}'$ and $\bar{W}' \neq 0$. Now E'' is measurable in M'' and $\bar{E}'' \geq \bar{W}'$ $(\bar{E}'' = \theta_{M'}(E''))$, so $s'(\bar{E}'') \geq s'(\bar{W}')$. It follows that $0 \geq r(\phi^{-1}(\bar{W}'))$, whence $\phi^{-1}(\bar{W}') = 0$, and $\bar{W}' = 0$, a contradiction.

Resumption of proof of Part B *of theorem.* Our next step is to set up a unitary transformation from $B_2(M)$ onto $B_2(M'')$. This transformation will be seen to be the unique extension to $B_2(M)$ of the contraction of $\tilde{\phi}$ to $\mathcal{B} \cap B_2(M)$.

Let f be an arbitrary element of $L_2(M) \cap L_\infty(M)$, and let $f'' = \phi(f)$. Since f'' is measurable with respect to M', and since every set measurable in M' is likewise measurable in M'', f'' is measurable with respect to M''. To evaluate the norm of f'' in $L_2(M'')$, we note that

$$\int_{R'} |f''(x)|^2 \, ds'(x) = -\int_0^\infty \lambda^2 \, ds'(S_\lambda'),$$

where $S_\lambda' = [x \mid |f''(x')| > \lambda]$. Now $\tilde{\phi}$ preserves order, on the subspace of real functions in $\mathcal{B}$ (where $p \geq q$ means that p-q is n. e. non-negative), for $p \geq q$, where p and q are real functions in $\mathcal{B}$, holds if and only if $p = q + r^2$, where r is a real function in $\mathcal{B}$; and this implies

$$\tilde{\phi}(p) = \tilde{\phi}(q) + (\tilde{\phi}(r))^2, \ \phi(p), \ \phi(q), \text{ and } \phi(r)$$

being real because ϕ preserves complex conjugation. Putting

$$S_\lambda = [x \mid |f(x)| > \lambda],$$

it is clear that $f\bar{f} \geq \lambda^2 \chi_{S_\lambda}$ and that $f''\bar{f}'' \geq \lambda^2 \chi_{S_\lambda}$. Applying $\tilde{\phi}$ to the first of these inequalities it results that $f''\bar{f}'' \geq \lambda^2 \chi_{\phi(S_\lambda)}$ which implies that $\phi(\bar{S}_\lambda) \leq \bar{S}_\lambda'$. Applying $\tilde{\phi}^{-1}$ to the second inequality yields the inequality $f\bar{f} \geq \lambda^2 \chi_{\phi^{-1}(S_\lambda')}$ which shows that $\phi^{-1}(\bar{S}_\lambda') \leq \bar{S}_\lambda$, and hence that $\bar{S}_\lambda' \leq \phi(\bar{S}_\lambda)$. It follows that $\phi(\bar{S}_\lambda) = \bar{S}_\lambda'$. Now $s'(\phi(\bar{E})) = r(\bar{E})$ for any $\bar{E} \varepsilon \mathcal{M}$, and hence

$$\int_{R'} |f''(x)|^2 \, ds'(x) = -\int_0^\infty \lambda^2 \, dr(S_\lambda) = \int_R |f(x)|^2 \, dr(x).$$

Thus $\tilde{\phi}$ maps $\mathfrak{B} \cap B_2(M)$ into $\mathfrak{B}' \cap B_2(M')$ in a manner which preserves the B_2 norm. Moreover, if f'' is any element of $\mathfrak{B}' \cap B_2(M')$, $\tilde{\phi}^{-1}(f'')$ is in $\mathfrak{B}$, and by the last line of the preceding paragraph (which is valid for arbitrary $f \in \mathfrak{B}$), it is also in $B_2(M)$. Thus $\tilde{\phi}$ maps $\mathfrak{B} \cap B_2(M)$ onto $\mathfrak{B}' \cap B_2(M')$. Since $\mathfrak{B} \cap B_2(M)$ and $\mathfrak{B}' \cap B_2(M')$ are dense in $B_2(M)$ and $B_2(M')$ respectively, we can conclude that there is a unique isometric linear map V of $B_2(M)$ onto $B_2(M')$ (i. e. V is unitary) such that $Vf = \tilde{\phi}(f)$, for $f \in \mathfrak{B} \cap B_2(M)$.

Next we construct the unitary map mentioned above, of $B_2(M'')$ onto $B_2(M')$. If E' is any σ-finite measurable set in M', it is easy to deduce from the Radon-Nikodym theorem (applied to the spaces whose set is E' and whose measurable sets are the M'-measurable sets of E', to which the same measures as those in M' and M'' are given) that there exists a measurable function $h_{E'}$ on M' which is n. e. positive on E', vanishes on the complement of E', and such that for any measurable subset F' of E',

$$s'(F') = \int_{F'} h_{E'}(x')\, dr'(x')\,; \quad r'(F') = \int_{F'} \{h_{E'}(x')\}^{-1}\, ds'(x').$$

We observe that if H' is any σ-finite measurable set in M' which contains E', then $h_{H'}(x') = h_{E'}(x')$ n. e. on E', for

$$\int_{F'} h_{H'}(x')\, ds'(x') = \int_{F'} h_{E'}(x')\, ds'(x'),$$

for any measurable subset F' of E'.

We now define a unitary map W on $L_2(M'')$ to $L_2(M')$ as follows: if $f \in L_2(M'')$, if E' is a σ-finite measurable set outside of which f vanishes, then $(Wf)(x') = f(x')\{h_{E'}(x')\}^{-\frac{1}{2}}$ for $x \in E'$ and $(Wf)(x') = 0$ if $x' \notin E'$. It is easy to verify that Wf is measurable on M', and (up to equivalence) is independent of the choice, within the stated restrictions, of E'; and $Wf \in L_2(M')$, in fact

$$\| Wf \|_2^2 = \int_{E'} |f(x')|^2 \{h_{E'}(x)\}^{-1} dr'(x') = \int_{E'} |f(x')|^2 ds'(x) = \| f \|_2^2$$

so that W is an isometry. To see that W is linear, let g be arbitrary in $L_2(M'')$ and let α and β be arbitrary constants. Then if F' is a σ-finite measurable set outside of which g vanishes, and if $G' = E' \cup F'$, we have

$$\begin{aligned}\{W(\alpha f + \beta g)\}(x') &= \{\alpha f(x') + \beta g(x')\}\{h_{G'}(x')\}^{-\frac{1}{2}} \\ &= \alpha f(x')\{h_{G'}(x')\}^{-\frac{1}{2}} + \beta g(x')\{h_{G'}(x')\}^{-\frac{1}{2}} \\ &= \alpha (Wf)(x') + \beta (Wg)(x').\end{aligned}$$

Thus W is unitary on $L_2(M'')$ to $L_2(M')$ and the product $U = WV$ is unitary on $L_2(M)$ to $L_2(M')$. It remains only to show that the algebraic isomorphism of $\mathcal{A}$ induced by U is actually Φ. That is, it must be shown that if $k \in \mathcal{B}$, then $UT_k = T_{\tilde{\phi}(k)}U$, where T_m denotes the operation of multiplication by the function m on the space of square-integrable functions over the measure space on which m is defined. Thus, the equation $UT_kf = T_{\tilde{\phi}(k)}Uf$ must be checked, for all $f \in L_2(M)$; but since U, T_k, and $T_{\tilde{\phi}(k)}$ are bounded operators, it is sufficient to check the equation for a dense set of f's, e. g. for $f \in L_2(M) \cap L_\infty(M)$. Now for such an f, $(UT_kf)(x') = (Up)(x')$, where $p(x) = k(x)f(x)$; $= (WVp)(x') = (Wq)(x')$, where $q = \tilde{\phi}(p) = \tilde{\phi}(k)\tilde{\phi}(f)$; $= q(x')\{h_{E'}(x')\}^{-\frac{1}{2}}$ on E' and 0 elsewhere, where E' is any σ-finite measurable set outside of which q vanishes; $= (\tilde{\phi}(k))(x')(\tilde{\phi}(f))(x')\{h_{E'}(x')\}^{-\frac{1}{2}}$ on E' and 0 elsewhere. On the other hand, $(T_{\tilde{\phi}(k)}Uf)(x') = (T_{\tilde{\phi}(k)}WVf)(x')$ $= (T_{\tilde{\phi}(k)}W\tilde{\phi}(f))(x') = (T_{\tilde{\phi}(k)}t)(x')$, where $t(x') = (\tilde{\phi}(f))(x')\{h_F'(x')\}^{-\frac{1}{2}}$ on F' and 0 elsewhere, F' being any σ-finite measurable set outside of which $\tilde{\phi}(f')$ vanishes; $= \tilde{\phi}(k)(x')t(x') = (\tilde{\phi}(k))(x')(\tilde{\phi}(f))(x')\{h_{F'}(x')\}^{-\frac{1}{2}}$ on F' and 0 elsewhere. Since $h_{E'}$ and $h_{F'}$ coincide n. e. on $E' \cap F'$, outside of which q vanishes, it results that $(UT_kf)(x') = T_{\tilde{\phi}(k)}Uf_{/}(x')$ n. e. and the proof is complete.

The next theorem relates weak equivalence of measure spaces to varieties of isomorphism of function-spaces over measure spaces. Such equivalence could also be related to unitary equivalence of the algebras of multiplications by bounded measurable functions vanishing outside of σ-finite sets, in a fashion similar to that described in the statement and proof of the preceding theorem.

THEOREM 4.2. *If the spaces X and X' of real-valued α-th-power integrable functions over two measure spaces M and M' are isomorphic, as partially ordered linear spaces ($\alpha \geqq 1$), then M and M' are weakly equivalent.*

If on the other hand M and M' are weakly equivalent, then there is an isometric isomorphism between X and X', as partially ordered Banach spaces.

We show first that M and M' are weakly equivalent if there is an isomorphism between X and X' as partially ordered linear spaces. Let ψ be such an isomorphism, i. e. an order-preserving linear map of X onto X'. For any numerical function f on a measure space, let $S(f)$ denote the set of points in the space where the function is nonzero. If $\bar{E}$ is any element of the σ-finite measure ring $\mathcal{F}$ of M, we define $\phi(\bar{E})$ as follows: if f is a non-negative element of X such that $\bar{S}(f)$ $(= \theta(S(f))) = \bar{E}$, then $\phi(\bar{E}) = \bar{S}(\psi(f))$. It is easy to see that such a function f always exists, so that $\phi(\bar{E})$ exists for all $\bar{E} \in \mathcal{F}$, and we show next that ϕ is single-valued.

Suppose that f and g are non-negative elements of X such that $\bar{S}(f) = \bar{S}(g) = \bar{E}$. Then setting $k = f \cup g$ and $h = f \cap g$ (where $\cup$ and $\cap$ denote the l. u. b. and g. l. b. respectively in X) it results that k and h are in X and that $\bar{S}(k) = \bar{S}(h) = \bar{E}$. Plainly, $\psi(k) \geq \psi(h)$, so that $\bar{S}(\psi(k)) \geq \bar{S}(\psi(h))$. On the other hand, if $\bar{S}(\psi(k)) \neq \bar{S}(\psi(h))$, then setting p' for a nonzero element of X' such that $\psi(k) \geq p'$ and $p' \cap \psi(h) = 0$ (e. g. $p'(x') = (\psi(k))(x')$ for $x' \in S(\psi(k)) - S(\psi(h))$ and $= 0$ otherwise), and $p = \psi^{-1}(p')$, it follows that $p \leq k$ and $p \cap h = 0$. These last relations imply the contradiction $p = 0$, and hence $\bar{S}(\psi(k)) = \bar{S}(\psi(h))$. Now from the inequalities $k \geq f \geq h$ it follows that $\bar{S}(\psi(k)) \geq \bar{S}(\psi(f)) \geq \bar{S}(\psi(h))$, and hence we can conclude that $\bar{S}(\psi(f)) = \bar{S}(\psi(k))$. Since k is a symmetric function of f and g, this equation implies that $\bar{S}(\psi(f)) = \bar{S}(\psi(g))$, i. e. ϕ is single-valued.

To complete the proof that M and M' are weakly equivalent it suffices to show that ϕ is onto, univalent, and order-preserving. If $\bar{E}'$ is in the σ-finite measure ring $\mathfrak{F}'$ of M', then there exists an element f' in X' such that $\bar{S}(f') = \bar{E}'$, and hence $\bar{E}' = \phi(\bar{E})$, where $\bar{E} = \bar{S}(\psi^{-1}(f'))$; i. e. ϕ is onto. To see that ϕ is univalent, let $\bar{F}$ and $\bar{G}$ be elements of $\mathfrak{F}$ such that $\phi(\bar{F}) = \phi(\bar{G})$. Putting f' for an element of X' for which $\bar{S}(f') = \phi(\bar{F}) = \phi(\bar{G})$, it follows from the definition of ϕ that if $f = \psi^{-1}(f')$, then $\bar{S}(f) = \bar{F} = \bar{G}$. Finally, if $\bar{F}$ and $\bar{G}$ are elements of $\mathfrak{F}$ such that $\bar{F} \leq \bar{G}$, then it is easily seen that there exist non-negative elements f and g of X for which $\bar{S}(f) = \bar{F}$, $\bar{S}(g) = \bar{G}$, and $f \leq g$. The last inequality implies that $\psi(f) \leq \psi(g)$, from which it results that $\bar{S}(\psi(f)) \leq \bar{S}(\psi(g))$, i. e. $\phi(\bar{F}) \leq \phi(\bar{G})$.

Now let ϕ be a weak equivalence between M and M', from $\mathfrak{F}$ onto $\mathfrak{F}'$. By Theorem 3. 4, M is metrically equivalent to a measure space P which is a direct sum of finite measure spaces, and hence localizable. It is easy to see from the proof of Theorem 3. 4 that there can be assumed to exist an isometric order-preserving linear map A of X onto $RL_\alpha(P)$. Since M is an arbitrary measure space, there is a map B of the same type as A on X' to $RL_\alpha(P')$, where P' is a localizable measure space. Since M and M' are weakly equivalent, so likewise are P and P', and hence, by Theorem 3. 3, P and P' are strongly equivalent.

It is not difficult to verify that the unitary transformation U of the proof of Theorem 4. 1 is order-preserving. Hence there exists an isometric linear order-preserving map U of $RL_2(P)$ onto $RL_2(P')$. It is easy to conclude that $B^{-1}UA$ is an isometric order-preserving linear map of X onto X', and so the proof of the theorem is complete in the case $\alpha = 2$. For general α the theorem can be concluded in a similar fashion as follows. Putting ξ for the algebraic isomorphism of $B_\infty(P)$ onto $B_\infty(P')$ induced by the strong

equivalence between P and P', and noting that in the present case $s' = r'$ (in the notation of the proof of Theorem 4. 1), it is straightforward to verify by the procedure of that proof, that the contraction of ξ to $B_\infty(P) \cap B_a(P)$ can be extended to an isometric, linear, order-preserving map U of $B_a(P)$ onto $B_a(P')$. Then, as above, $B^{-1}UA$ is the required map of X onto X'.

We conclude this section by characterizing metric equivalence in terms of the induced correspondence between certain function spaces.

THEOREM 4. 3. *Two measure spaces M and M' are metrically equivalent if and only if there is an algebraic isomorphism between the algebras of bounded real-valued measurable functions vanishing at infinity on M and M' which preserves the integrals of idempotents in the algebras.*

Let $\mathcal{A}$ and $\mathcal{A}'$ be the algebras of bounded real-valued measurable functions vanishing at infinity on $M = (R, \mathcal{R}, r)$ and $M' = (R', \mathcal{R}', r')$, and let ϕ be an algebraic isomorphism from $\mathcal{A}$ onto $\mathcal{A}'$ which preserves the integrals of idempotents in $\mathcal{A}$. It is easy to see that an idempotent f in $\mathcal{A}$ is the characteristic function of a chunk in M, and that conversely every such function is an idempotent in $\mathcal{A}$. It follows that if E is any chunk in M, there is a chunk E' in M' such that $\phi(\chi_{\bar{E}}) = \chi_{\bar{E}'}$, and $r(E) = r'(E')$. Now let $\mathcal{D}$ and $\mathcal{D}'$ be the subringc of $\mathcal{F}$ and $\mathcal{F}'$ respectively which consist of the elements of finite measure, and put $\tilde{\phi}$ for the mapping on $\mathcal{D}$ to $\mathcal{D}'$ determined by the equation $\tilde{\phi}(\bar{E}) = \bar{E}'$, where $\phi(\chi_{\bar{E}}) = \chi_{\bar{E}'}$, for $\bar{E} \in \mathcal{D}$. It is easy to see that ϕ is a measure-preserving ring isomorphism of $\mathcal{D}$ onto $\mathcal{D}'$.

Now let M and M' be metrically equivalent, and let η be a measure-preserving ring isomorphism of the finite measure ring $\mathcal{D}$ of M onto the corresponding ring $\mathcal{D}'$ of M'. Put $\mathcal{L}$ and $\mathcal{L}'$ for the subalgebras of $\mathcal{A}$ and $\mathcal{A}'$ consisting of all real finite linear combinations of characteristic functions of chunks in M and M' respectively. It is easy to see by the procedure used in the proof of Part A of Theorem 4. 1, that the mapping $\phi_0 : \Sigma_i \alpha_i \chi_{\bar{E}_i} \to \Sigma_i \alpha_i \chi_{\eta(\bar{E}_i)}$, where $\bar{E}_i \in \mathcal{D}$, i ranges over a finite set, and the α_i are real, is a (single-valued) algebraic isomorphism of $\mathcal{L}$ onto $\mathcal{L}'$. Moreover, $\| \phi_0(f) \|_\infty \leq \| f \|_\infty$, for $f \in \mathcal{L}$. It is not difficult to verify that $\mathcal{L}$ and $\mathcal{L}'$ are dense in $\mathcal{A}$ and $\mathcal{A}'$ respectively (with respect to the B_∞ norm topology), and it follows that ϕ_0 can be extended to an algebraic isomorphism of $\mathcal{A}$ onto $\mathcal{A}'$. It is obvious that ϕ preserves the integrals of idempotents in $\mathcal{A}$.

We need the following definition before stating our final characterization of metric equivalence.

DEFINITION 4. 2. Let M and M' be measure spaces. A mapping ϕ on $L_2(M)$ to $L_2(M')$ is called *multiplicative* if whenever elements f, g, and h

of $L_2(M)$ are such that $f(x)g(x) = h(x)$ n. e. on M, then the equation $f'(x')g'(x') = h'(x')$ holds n. e. on M', where f', g', and h' are the transforms of f, g, and h respectively, under ϕ.

THEOREM 4. 4. *Two measure spaces M and M' are metrically equivalent if and only if there is a multiplicative unitary transformation from $L_2(M)$ onto $L_2(M')$.*

We treat first the "only if" part of the theorem. Let η be a measure-preserving ring isomorphism of the finite measure ring $\mathfrak{D}$ of $M = (R, \mathfrak{R}, r)$ onto the corresponding ring $\mathfrak{D}'$ of M', and let $\mathfrak{L}$ and $\mathfrak{L}'$ be the algebras of all complex finite linear combinations of characteristic functions of chunks in M and M' respectively. Then as in the proof of Theorem 4. 1, the mapping: $\phi_0 : \Sigma_i \alpha_i \chi_{\bar{E}_i} \to \Sigma_i \alpha_i \chi_{\eta(\bar{E}_i)}$, where $\bar{E}_i \in \mathfrak{D}$, i ranges over a finite set, and the α_i are complex numbers, is a (single-valued) algebraic isomorphism of $\mathfrak{L}$ onto $\mathfrak{L}'$. Just as in the proof of the preceding theorem ϕ_0 can be extended to an algebraic isomorphism ϕ of the algebra $\mathfrak{C}$ of all complex-valued measurable functions vanishing at infinity on M, onto the corresponding algebra $\mathfrak{C}'$ on M' (both algebras modulo the ideals of functions equivalent to zero). Plainly $\mathfrak{C} \cap B_2(M)$ and $\mathfrak{C}' \cap B_2(M')$ are dense subsets of $B_2(M)$ and $B_2(M')$ respectively, and it can be verified as in the proof of Theorem 4. 1, that ϕ is isometric, relative to the B_2 norm, from $\mathfrak{C} \cap B_2(M)$ to $\mathfrak{C}' \cap B_2(M')$. It follows that the contraction of ϕ to $\mathfrak{C} \cap B_2(M)$ can be uniquely extended to a unitary transformation U from $B_2(M)$ onto $B_2(M')$.

To complete the proof of this part of the theorem, it suffices to show that U is multiplicative. Let f, g, and h be elements of $L_2(M)$ such that $f(x)g(x) = h(x)$ n. e. on M, and set

$$F_n = [x \mid |f(x)| \leq n], \quad G_n = [x \mid |g(x)| \leq n], \quad H_n = [x \mid |h(x) \leq n],$$
$$\text{and } K_n = F_n \cap G_n \cap H_n \qquad (n = 1, 2, \cdots).$$

Plainly each of $\bigcup_n F_n$, $\bigcup_n G_n$, $\bigcup_n H_n$ is equivalent to R, and hence so also is $\bigcup_n K_n$. Now let $f_n = f\chi_{K_n}$, $g_n = g\chi_{K_n}$, and $h_n = h\chi_{K_n}$. Clearly, $f_n \to f$, $g_n \to g$, and $h_n \to h$, in $L_2(M)$, as $n \to \infty$; $f_n(x)g_n(x) = h_n(x)$ n. e. on M; and f_n, g_n, and h_n are in $\mathfrak{C}$. It results in particular that $Uf_n = \phi(f_n) = f_n'$, $Ug_n = \phi(g_n) = g_n'$, and $Uh_n = \phi(h_n) = h_n'$. Now let $\{n_i\}$ be a subsequence of the positive integers such that $f_{n_i}'(x') \to f'(x')$, $g_{n_i}'(x') \to g'(x')$, and $h_{n_i}'(x') \to h'(x')$ n. e. on M', where $f' = Uf$, $g' = Ug$, and $h' = Uh$. Observing that $h_n' = \phi(f_n g_n) = \phi(f_n)\phi(g_n) = f_n' g_n'$, we have n. e. on

$$M' h'(x') = \lim_i h_{n_i}'(x') = \lim_i f_{n_i}'(x') g_{n_i}'(x')$$
$$= \lim_i f_{n_i}'(x') \lim_i g_{n_i}'(x') = f'(x') g'(x').$$

Now conversely let T be a multiplicative unitary transformation from $L_2(M)$ onto $L_2(M')$. If E is a chunk in M, then putting k for the characteristic function of E, $T(k^2) = T(k) = (T(k))^2$, so that $T(k)$ is a characteristic function of a measurable set E', and as T is unitary, $\| k \|_2 = \| k' \|_2$, where k' is the characteristic function of E', whence $r(E) = r'(E')$. Putting $\bar{E}' = \psi(\bar{E})$, it follows readily by a device previously employed that ψ is order-preserving, and ψ is clearly one-to-one and onto the finite measure ring of M'. It follows that ψ is an algebraic isomorphism of the finite measure ring of M onto that of M', and hence is a metric equivalence of M with M'.

5. Localizable spaces. We show in the present section that localizable spaces are characterized by any of a number of important measure-theoretic properties. These results, together with those of the preceding sections, suggests that, at least as regards applications to many aspects of functional analysis, the localizable spaces constitute the most significant general class of measure spaces properly containing the finite ones, and are of the maximum generality consistent with usefulness in any large sector of abstract analysis.

Before stating the main theorem of this section we make the following

DEFINITIONS 5.1. A collection $\mathcal{K}$ of chunks in a measure space [or of elements of the finite measure ring $\mathcal{D}$ of the space] is called a *conditional σ-ideal* if it is an ideal in the ring of measurable sets [or in $\mathfrak{M}$], and if any chunk [or element of $\mathcal{D}$] which is a countable union of elements of $\mathcal{K}$ is itself in $\mathcal{K}$. An ideal $\mathcal{L}$ of chunks [or of elements of $\mathcal{D}$] is said to be *localized* in L [or $\bar{L}$] if there is a measurable set L [or element $\bar{L}$ of the measure ring] such that every element of $\mathcal{L}$ is equivalent to a subset of L, and conversely every chunk contained in L is equivalent to an element of $\mathcal{L}$ [or such that $\mathcal{L} = [\bar{L} \cap \bar{D} \mid \bar{D} \in \mathcal{D}]$].

DEFINITIONS 5.2. The *Radon-Nikodym theorem is* said to be *vaild for a space* $M = (R, \mathfrak{R}, r)$ if whenever $M' = (R, \mathfrak{R}, r')$ is a measure space such that $r'(E) \leq \alpha r(E)$ for $E \in \mathfrak{R}$, α being a constant, then there exists a measurable function f on M such that for $E \in \mathfrak{R}$, $r'(E) = \int_E f(x)\,dr(x)$. The *Riesz representation theorem* for linear functionals on L_1 *is* said to be *valid* for M if for every continuous linear functional ϕ on $L_1(M)$ there is a (fixed) bounded measurable function k on M such that

$$\phi(f) = \int_R k(x)f(x)\,dr(x) \quad (f \in L_1(M)).$$

Remarks 5.1. It is easy to see that if a collection $\mathcal{K}$ of chunks is a

conditional σ-ideal, then the same is true of $\theta(\mathcal{K})$; and that $\mathcal{K}$ is localized in k if and only if $\theta(\mathcal{K})$ is localized in $\theta(K)$.

The form of the Radon-Nikodym theorem given above differs somewhat from the usual version. It is easy to see however by a well-known device that the result for the case when r' is merely absolutely continuous relative to r follows from that for the special case described.

The following theorem further justifies our term "localizable," by showing that such spaces are precisely those in which every conditional σ-ideal of chunks is localized.

THEOREM 5.1. *Any of the following conditions on a measure space $M = (R, \mathfrak{R}, r)$ implies all of the other conditions.*

a. M is localizable.

b. M is strongly equivalent to a direct sum of finite measure spaces.

c. The Radon-Nikodym theorem is valid for M.

d. The multiplication algebra of M is maximal abelian in the algebra of all bounded linear operators on $L_2(M)$.

e. Every conditional σ-ideal of chunks is localized.

f. The Riesz representation theorem for linear functionals on L_1 is valid for M.

That a and b are equivalent was proved in Section **3** and is stated here for convenience of reference. In order to prove the theorem, it therefore suffices to show that a and e are equivalent and to establish the following cycle of implications: $e \Rightarrow d$, $d \Rightarrow f$, $f \Rightarrow c$, and $c \Rightarrow e$.

A: $a \Rightarrow e$. Let $\mathcal{K}$ be a conditional σ-ideal in the finite measure ring $\mathcal{D}$ of the localizable space $M = (R, \mathfrak{R}, r)$. Setting $\bar{K} = \text{l. u. b.}_{\bar{E} \in \mathcal{K}} \bar{E}$, then every element of $\mathcal{K}$ is bounded by $\bar{K}$, and it remains to show that if $\bar{F} \leq \bar{K}$ and $\bar{F} \in \mathcal{D}$, then $\bar{F} \in \mathcal{K}$. Now for such an $\bar{F}$ we have

$$\bar{F} = \bar{F} \cap \bar{K} = \bar{F} \cap \text{l. u. b.}_{\bar{E} \in \mathcal{K}} \bar{E} = \text{l. u. b.}_{\bar{E} \in \mathcal{K}} (\bar{E} \cap \bar{F}).$$

Choosing $\bar{E}_i$ in $\mathcal{K}$ such that $r(\bar{E}_i \cap \bar{F}) \to \text{l. u. b.}_{\bar{E} \in \mathcal{K}} r(\bar{E} \cap \bar{F})$, it is easy to verify that $\bar{F} \leq \bigcup_i \bar{E}_i$, and hence $\bar{F} \in \mathcal{K}$.

B: $e \Rightarrow a$. Let $\mathcal{L}$ be any subset of the measure ring $\mathfrak{M}$ of the measure space $M = (R, \mathfrak{R}, r)$, for which e holds; it must be shown that $\mathcal{L}$ has a l. u. b. in $\mathfrak{M}$. Set $\mathcal{L}_0$ for the set of all elements of $\mathcal{D}$, the finite measure ring of $\mathfrak{M}$, which are bounded by some element of $\mathcal{L}$, and let $\mathcal{K}$ be the conditional σ-ideal generated by $\mathcal{L}_0$. On the one hand $\mathcal{K}$ is the intersection of all conditional σ-ideals in $\mathcal{D}$ which contain $\mathcal{L}_0$ and so exists; and on the

other hand, it is easy to verify that $\mathcal{K}$ consists of all elements of $\mathcal{D}$ which are countable unions of elements of $\mathcal{D}$ which are bounded by elements of $\mathcal{L}$. Now let $\mathcal{K}$ be localized in $\bar{K}$. Then $\bar{K}$ is an upper bound for the elements of $\mathcal{L}$, since it is an upper bound for the elements of $\mathcal{D}$ which are bounded by the elements of $\mathcal{L}$, and since by Lemma 3. 3. 1 any element of $\mathcal{L}$ is the l. u. b. of the elements of $\mathcal{D}$ which it bounds. Furthermore, $\bar{K}$ is the l. u. b., for if $\bar{K}'$ is any other upper bound, it is clear from the form of $\mathcal{K}$ that $\bar{K}'$ is an upper bound for $\mathcal{K}$. This implies that $\bar{K} - \bar{K}'$ is disjoint from every element of $\mathcal{K}$, but $(\bar{K} - \bar{K}') \cap \bar{F}$ is in $\mathcal{K}$ for every $\bar{F} \in \mathcal{D}$, being bounded by $\bar{K}$, and hence $(\bar{K} - \bar{K}') \cap \bar{F} = 0$ for all such $\bar{F}$, whence $\bar{K} - \bar{K}' = 0$, i. e. $\bar{K}' \geq \bar{K}$.

C: $e \Rightarrow d$. If M is a measure space for which e holds, then as just proved, M is localizable, and hence by Theorem 3. 4 strongly equivalent to a direct sum M' of finite measure spaces. By Theorem 4. 1 the multiplication algebra of M is unitarily equivalent to the multiplication algebra of M', and hence it suffices to show that d holds for any direct sum of finite measure spaces, say for $M = \sum_{\xi \in \Xi} \oplus M_\xi$, where $M_\xi = (R_\xi, \mathcal{R}_\xi, r_\xi)$ is finite.

Let U be a bounded linear operator on $L_2(M)$ which commutes with all multiplications T_k by bounded measurable functions k on M. Let χ_ξ be the characteristic function of R_ξ and let P_ξ be the operation of multiplication by χ_ξ. Then $T_k U \chi_\xi = U T_k \chi_\xi$, for $\xi \in \Xi$ and $k \in L_\infty(M)$, and putting $k = \chi_\eta$ it follows that $P_\eta U \chi_\xi = U \chi_\xi$, i. e. $U\chi_\xi$ vanishes n. e. outside of R_η. Now if g is any element of $L_2(M) \cap L_\infty(M)$ which vanishes outside of R_η, then $Ug = UT_g \chi_\xi = T_g U \chi_\xi = U_\xi g$, where U_ξ is the operation of multiplication by $U\chi_\xi$ on the domain $L_2(M) \cap L_\infty(M)$. Clearly $\| U_\xi g \| = \| Ug \| \leq \| U \| \| g \|$ for such g, and in particular, putting g for the characteristic function of the subset Z of R_ξ on which $U\chi_\xi$ exceeds $\| U \| + 1$ in absolute value, we have $(\| U \| + 1)(r(Z))^{\frac{1}{2}} \leq \| U_\xi g \| \leq \| U \| (r(Z))^{\frac{1}{2}}$. It follows that $r(Z) = 0$, which implies that $U\chi_\xi$ is bounded n. e. by $\| U \| + 1$ on R_ξ.

Now let h be the function on R which coincides on R_ξ with $U\chi_\xi$. It is plain that h is measurable, essentially bounded by $\| U \| + 1$, and that if g is in $L_2(M)$ and vanishes outside of some R_ξ, then $Ug = T_h g$. The last equation holds also if g vanishes outside of the union of a finite number of the R_ξ, and thus holds for a set of g's which is dense in $L_2(M)$. Since U and T_h are bounded operators, it results that they are the same.

D: $d \Rightarrow f$. The following proof is related to the proofs of the Radon-Nikodym theorem which have been given by von Neumann [8] and Stone [6]. Let d hold for the measure space $M = (R, \mathcal{R}, r)$ and let ϕ be a continuous

linear functional on $L_1(M)$ of bound α. We define a hermitian bilinear form Φ on $L_2(M)$ as follows: if $f, g \in L_2(M)$, $\Phi(f, g) = \phi(f\bar{g})$. Then $|\Phi(f, g)| \leq \alpha \| f\bar{g} \|_1 \leq \alpha \| f \|_2 \| g \|_2$, so that for each fixed g, $\Phi(f, g)$ is a bounded linear functional of f, of bound not greater than $\alpha \| g \|_2$. By the representation theorem for bounded linear functionals on Hilbert space, there exists an element $H(g)$ in $L_2(M)$ such that $\Phi(f, g) = (f, H(g))$, and $\| H(g) \|_2 \leq \alpha \| g \|_2$. It follows easily that H is a bounded linear operator on $L_2(M)$.

Now H^* commutes with all operators on $L_2(M)$ of the form T_k, where T_k is multiplication by the bounded measurable function k, for if f and g are arbitrary in $L_2(M)$, then $(H^*T_k f, g) = (T_k f, Hg) = \Phi(T_k f, g) = \phi(kf\bar{g}) = \Phi(f, T_{\bar{k}}g) = (f, HT_{\bar{k}}g)$, and noting that $(T_k)^* = T_{\bar{k}}$, the last expression is $(T_k H^* f, g)$. Since d holds for M, $H = T_p$ for some p in $L_\infty(M)$. It results now from the fact that the product of two arbitrary functions in $L_2(M)$ is an arbitrary function in $L_1(M)$ that for all $f \in L_1(M)$, $\phi(f) = \int_R p(x) f(x)\, dr(x)$.

E: $f \Rightarrow c$. Now let f hold in the measure space $M = (R, \mathfrak{R}, r)$ and let $M' = (R, \mathfrak{R}, r')$ be a measure space for which $r'(E) \leqq \alpha r(E)$, $E \in \mathfrak{R}$, α being a fixed positive number. We define a functional ϕ on $L_1(M)$ by the equatioin $\phi(f) = \int_R f(x)\, dr'(x)$, $f \in L_1(M)$; the indicated integral exists, and in fact

$$\int_R |f(x)|\, dr'(x) = \int_0^\infty \lambda\, dr'[x \mid |f(x)| \leq \lambda]$$

$$\leq \alpha \int_0^\infty \lambda\, dr[x \mid |f(x)| \leq \lambda] = \alpha \int_R |f(x)|\, dr(x),$$

i.e. $|\phi(f)| \leq \alpha \| f \|_1$. As f holds, there exists $p \in L_\infty(M)$ such that $\phi(f) = \int_R p(x) f(x)\, dr(x)$ for $f \in L_1(M)$, and it is readily shown that $0 \leq p(x) \leq \alpha$ n. e. on M. Thus for $f \in L_1(M)$,

$$\int_R f(x)\, dr'(x) = \int_R p(x) f(x)\, dr(x),$$

and putting f equal to the characteristic function of a set in $\mathfrak{R}$, c follows.

F: $c \Rightarrow e$. Let $M = (R, \mathfrak{R}, r)$ satisfy c, and let $\mathfrak{K}$ be a conditional σ-ideal of chunks in M. We show that $\mathfrak{K}$ is localized essentially by constructing a kind of quotient space of M with respect to $\mathfrak{K}$, and deriving from this quotient a measure which is absolutely continuous with respect to r.

If $E \in \mathcal{R}$, the least upper bound of $\bar{E} \cap \bar{F}$ as $\bar{F}$ varies over $\theta(\mathcal{K})$ exists and is in $\theta(\mathcal{K})$, for if $\bar{F}_i$ $(i = 1, 2, \cdots)$ is a sequence of elements of $\mathcal{K}$ for which $\lim_i r(\bar{E} \cap \bar{F}_i) = \text{l. u. b.}_{F \in \theta(\mathcal{K})} r(\bar{E} \cap \bar{F})$, then it is easily seen that $\bar{E} \cap \bigcup_i \bar{F}_i$ is the least upper bound in question, and it is clear from the circumstance that $\theta(\mathcal{K})$ is a conditional σ-ideal that it is in $\theta(\mathcal{K})$. We now put $\psi(\bar{E})$ for that least upper bound and define a function r' on R as follows: $r'(E) = r(\psi(\bar{E}))$, $E \in \mathcal{R}$. Evidently $r'(E) \leq r(E)$, for $E \in \mathcal{R}$, and we show next, employing a simplification of our original argument due to the referee, that r' is a countably additive measure on $\mathcal{R}$.

Let $\{E_i\}$ be a sequence of mutually disjoint elements of $\mathcal{R}$ whose union E is in $\mathcal{R}$. Then $\psi(\bar{E}) = \psi(\bar{E}) \cap \bar{E} = \bigcup_i (\psi(\bar{E}) \cap \bar{E}_i) \leq \bigcup_i \psi(\bar{E}_i)$ (for $\psi(\bar{E}) \in \theta(\mathcal{K})$). On the other hand, it is plain that $\bigcup_i \psi(\bar{E}_i) \leq \psi(\bar{E})$, so $\psi(\bar{E}) = \bigcup_i \psi(\bar{E}_i)$. Hence $r'(E) = r(\psi(E)) = \Sigma_i r(\psi(\bar{E}_i))$ (as $\psi(\bar{E}_i) \leq \bar{E}_i$, the $\psi(\bar{E}_i)$ are mutually disjoint) $= \Sigma_i r'(E_i)$. By the Radon-Nikodym theorem, $r'(E) = \int_E f(x)\,dr(x)$ for $E \in \mathcal{R}$, where f is a measurable function with values between 0 and 1. It is readily verified that $\psi(\psi(\bar{E})) = \psi(\bar{E})$, so $r(\psi(\psi(\bar{E}) = r(\psi(\bar{E}))$, whence $\int_{\psi(E)} f(x)\,dr(x) = r(\psi(E))$, where $\psi(E)$ is any set in the residue class $\psi(\bar{E})$. As $f(x) \geq 0$, it results that $f(x) = 1$ n. e. on $\psi(E)$, and as $r'(E - \psi(E)) = 0$, $f(x) = 0$ n. e. on $E - \psi(E)$. Now E is a general element of $\mathcal{R}$, so f is n. e. equal either to zero or to one, and hence is equivalent to the characteristic function of a measurable set K. It follows without difficulty that $\mathcal{K}$ is localized in K.

We close this section by giving a useful sufficient condition for a space to be localizable. For definitions and results concerning uniform spaces we refer to [9].

THEOREM 5.2. *Any uniformly locally compact regular measure space is the direct sum of finite measure spaces* (*and hence localizable*).

Let $M = (R, \mathcal{R}, r)$ be a uniformly locally compact regular measure space, and let $N(.)$ be a uniform family of neighborhoods in R such that $N = N^{-1}$ and $N^2(x) \subset P(x)$ for $x \in R$, where $P(.)$ is a neighborhood family such that $P(x)$ is compact for $x \in R$. Then the sets $Q(x) = \bigcup_{i=1}^{\infty} N^i(x)$ are σ-compact, and open and closed, and it is easily seen that $Q(x)$ meets $Q(x')$ if and only if $Q(x) = Q(x')$. Hence R is the union of mutually disjoint σ-compact open and closed sets, say R_ξ, with ξ ranging over the index set Ξ.

Now any set in $\mathcal{R}$ is in the σ-ring determined by the compact subsets of R, and any such compact subset meets at most finitely many of the R_ξ

Hence each element of $\mathfrak{R}$ meets at most countably many of the R_ξ. It follows that M is the direct sum over Ξ of the measure spaces $M_\xi = (R_\xi, \mathfrak{R}_\xi, r)$, where $\mathfrak{R}_\xi = [E \cap R_\xi \mid E \in \mathfrak{R}]$. Each M_ξ is clearly σ-compact and a regular locally compact measure space. Therefore M_ξ is σ-finite, and hence a direct sum of countably many finite measure spaces. It results that M is itself a direct sum of finite measure spaces.

COROLLARY 5. 2. *Any locally compact group is localizable, relative to Haar measure (on the conditional σ-ring determined by the compact subsets).*

Remark 5. 2. The following question arises fairly naturally in connection with this section: is any measure space strongly equivalent to a complete measure space (i. e. one in which every subset of a set equivalent to zero is also equivalent to zero)? Any measure space has a "completion" and it is easily seen that a localizable space is strongly equivalent to its completion, but the question appears to be open in general. An affirmative answer would be useful, but it seems plausible that the answer is negative.

We conclude this section by giving a partial extension of a theorem of Dieudonné [1], and also a short proof of the original theorem. As a matter of terminology, if $\mathfrak{N}$ is a subring of the measure ring of a measure space, a function on the space will be said to be measurable relative to $\mathfrak{N}$ if it is measurable in the usual sense relative to $\theta^{-1}(\mathfrak{N})$.

THEOREM 5. 3. *Let $M = (R, \mathfrak{R}, r)$ be a localizable measure space with measure ring $\mathfrak{M}$ and let $\mathfrak{N}$ be a complete subring of $\mathfrak{M}$ containing $\bar{R}$. Let $\mathfrak{A}$ be the algebra of all bounded measurable functions on M that are also measurable relative to $\mathfrak{N}$. Then*

a) *there exists a localizable measure space $N = (S, \mathfrak{S}, s)$ and an algebraic isomorphism ψ of $\mathfrak{A}$ onto the algebra of all bounded measurable functions on N (so that in particular $\mathfrak{N}$ is algebraically isomorphic to the measure ring of N);*

b) *there exists a positive continuous linear map ϕ of $L_1(M)$ to $L_1(N)$ such that for $f \in L_1(M)$ and $g \in \mathfrak{A}$,*

$$\int_M fg = \int_N \phi(f)\psi(g).$$

We mention that the measure space N is not unique within isomorphism or metric equivalence, tho necessarily unique within strong equivalence. In the special case (treated by Dieudonné) that M is finite there is a natural way of normalizing N, which has the useful feature that $\phi(f)$ is bounded when f is, but in general no such normalization appears possible.

The following lemma is somewhat more general than is strictly necessary for the proof of the theorem, but it has some independent interest.

LEMMA 6.3.1. *A complete subring of a complete measure ring is algebraically isomorphic to the measure ring of a measure space.*

If $(\mathfrak{M}, m)$ is a complete measure ring, then $\mathfrak{M}$ itself is isometrically algebraically isomorphic with the measure ring of a measure space. For the unit element I of $\mathfrak{M}$ is the l. u. b. of a collection of elements on which m is finite, from which it follows readily that it is the l. u. b. of a collection of mutually disjoint elements $\{A_\mu\}$ on each of which m is finite. The finite measure ring $(A_\mu \cap \mathfrak{M}, m)$ is well-known to be isometrically algebraically isomorphic with the measure ring of a measure space M_μ, and it follows without difficulty that $(\mathfrak{M}, m)$ is isometrically alegbraically isomorphic with the measure ring of the direct sum of the M_μ.

Now let $\mathfrak{N}$ be a complete subring of $\mathfrak{M}$, which we now take to be the measure ring of the measure space M. If I' is the l. u. b. of the elements of $\mathfrak{N}$, then $\mathfrak{N}$ is also a complete subring of the complete measure ring $(I' \cap \mathfrak{M}, m)$, so it is no essential loss of generality to assume that $I \in \mathfrak{N}$.

Let an element P of $\mathfrak{N}$ be called *finite* if there exists a nonzero element E in $\mathfrak{M}$ of finite measure such that when $A, B \in \mathfrak{N}$, $A, B \leq P$, and $A \cap E = B \cap E$, then $A = B$; we call E a *separating element* for P. Let $\mathfrak{F}$ be a family of nonzero finite elements of $\mathfrak{N}$ which is maximal with respect to mutual disjointness. Set l. u. b.$_{A \in \mathfrak{F}} A = Q$; we show next that $Q = I$. For otherwise there exists a nonzero element G of finite measure in $\mathfrak{M}$ such that $I - Q \geq G$. Putting R_0 for the l. u. b. of the elements R in $\mathfrak{N}$ for which $R \leq I - Q$ and $R \cap G = 0$, then $I - Q - R_0$ is disjoint from all the elements of $\mathfrak{F}$, is nonzero as $G \neq 0$, and is finite, since if $A, B \in \mathfrak{N}$, $A, B \leq I - Q - R_0$, and $A \cap G = B \cap G$, then $(A \ominus B) \cap G = 0$, while $A \ominus B \leq I - Q$, which by the definition of R_0 implies that $A \ominus B \leq R_0$. This inequality combined with the inequality $A, B \leq I - Q - R_0$, shows that $A \ominus B = 0$, or $A = B$.

Now put n for the function on $\mathfrak{N}$ defined by the equation

$$n(A) = \sum_{P \in \mathfrak{F}} m(\sigma(P) \cap A),$$

where $\sigma(P)$ is a separating element for P. Then it is clear that n is a countably additive non-negative-valued function, and to complete the proof it need only be shown that each element A of $\mathfrak{N}$ is the l. u. b. of elements on which n is finite. But it is clear that $[\sigma(P) \cap A \mid P \in \mathfrak{F}]$ is such a collection.

Proof of theorem. Let $N = (S, \mathcal{S}, s)$ be a (localizable) measure space whose measure ring is algebraically isomorphic to $\mathcal{N}$, and let $\mathcal{A}'$ be the algebra of all bounded measurable functions on N. Now every element in $\mathcal{A}'$ is the uniform limit of a sequence of finite linear combinations of characteristic functions of mutually disjoint measurable sets, and every such linear combination corresponds in an obvious way with a linear combination of characteristic functions of mutually disjoint elements of $\mathcal{N}$. As this correspondence is an isometry, it can be uniquely extended to an algebraic isomorphism between $\mathcal{A}'$ and the uniform closure of the set of all bounded finite linear combinations of characteristic functions of mutually disjoint elements of $\mathcal{N}$, which is simply $\mathcal{A}$. We put ψ for this isomorphism, as a map on $\mathcal{A}$ to $\mathcal{A}'$

Now let f be arbitrary in $L_1(M)$, and for an arbitrary measurable set E in N, put $\eta(E)$ for the element of $\mathcal{N}$ such that $\psi(\chi_{\eta(E)}) = \chi_E$. Then setting $m_f(E) = \int_{\eta(E)} f$, m_f is countably additive and is absolutely continuous with respect to s. As N is localizable, there exists a measurable function F on N such that $m_f(E) = \int_E F$ (altho m_f need not be bounded relative to s, the conclusion of the Radon-Nikodym theorem can be shown without difficulty to remain valid). Clearly F is non-negative when f is, and $\int_N F = \int_M f$, so F is integrable. Thus in case g is the characteristic function of a set in $\mathcal{N}$, we have $\int_M fg = \int_N \phi(f)\psi(g)$, if we set $F = \phi(f)$. It follows that this formula holds also for all finite linear combinations of such g's, hence for all finite linear combinations of characteristic functions of mutually disjoint sets in $\mathcal{N}$, and finally for all $g \in \mathcal{A}$.

THEOREM 5.4. (Dieudonné). *Let $M = (R, \mathcal{R}, r)$ be a finite perfect measure space and $\mathcal{T}$ a σ-subring of $\mathcal{R}$ containing R. Then there exists a perfect measure space $N = (S, \mathcal{S}, s)$, a continuous map ϕ of R onto S, and a positive linear map ψ of $\mathcal{B}(R)$ onto $\mathcal{B}(S)$ such that* a) *an element g of $\mathcal{B}(R)$ is (equivalent to) a function measurable relative to $\mathcal{T}$ if and only if it has the form $g = h \circ \phi$, for some $h \in \mathcal{B}(S)$, and for any E in $\mathcal{S}$, $s(E) = r(\phi^{-1}(E))$;* b) *if $f \in \mathcal{B}(R)$ and $h \in C(S)$, then*

$$\int_M f(h \circ \phi) = \int_N \psi(f)h.$$

The "$\circ$" in the statement of the theorem signifies functional composition. We mention that if $m_y (y \in S)$ is the regular measure on R such

that $\psi(f)(y) = \int f(x)dm_y(x)$ (by the Riesz-Markoff theorem, such a measure exists), then m_y is concentrated on $\phi^{-1}(y)$, as Dieudonné shows by a brief argument.

To prove the theorem, let $\mathcal{A}$ be the algebra of all continuous functions on R which are equivalent to functions measurable relative to $\mathfrak{I}$. By well-known results, $\mathcal{A}$ is algebraically isomorphic to $\mathcal{B}(S)$, for some compact Hausdorff space S. Now $\mathcal{A}$ is a closed subring of $\mathcal{B}(R)$, so by a theorem of Stone [7], there exists a continuous map ϕ of R onto S such that $\mathcal{A}$ consists of all functions of the form $h \circ \phi$, with $h \in \mathcal{B}(S)$. Putting $s(E) = r(\phi^{-1}(E))$ for $E \in \mathfrak{S}$, where $\mathfrak{S}$ denotes the collection of all Borel sets in S, then $N = (S, \mathfrak{S}, s)$ is a regular measure space, and can be shown without difficulty to be perfect (cf. e. g. the proof of Theorem 6. 1 below). If $f \in \mathcal{B}(M)$, and $m_f(E) = \int_{\phi^{-1}(E)} f \; (E \in \mathfrak{S})$, then $|m_f| \leq \|f\| s(E)$, so by the Radon-Nikodym theorem there exists a bounded measurable function F such that $m_f(E) = \int_E F$, and as N is perfect, F can be taken to be in $\mathcal{B}(S)$. Putting $\psi(f) = F$, it follows by the argument used at the end of the proof of the last theorem that conclusion b) of the present theorem holds, and it has already been shown that a) holds.

6. Perfect spaces. In various situations it is desirable to replace a given measure space by a metrically equivalent one in which measurability difficulties are minimized, even tho the new space may, roughly sepaking, be much larger than the original one. In this section we show how a given space may be altered, or "perfected," in this way, and that this can be done in essentially only one way. The existence of a perfection of a finite measure space was shown by Kakutani, but his technique apparently does not extend to the general case. The uniqueness referred to in the following theorem is with respect to the following concept of identity of regular measure spaces: two such spaces are abstractly identical (as regular measure spaces) if they are isomorphic as measure spaces via an isomorphism which is also a topological homeomorphism.

THEOREM 6. 1. *Any measure space is metrically equivalent to a unique perfect space.*

Having shown that any measure space is metrically equivalent to a localizable space, it suffices to show that any localizable space is metrically equivalent to a unique perfect space. Let $M = (S, \mathfrak{S}, s)$ be an arbitrary

localizable measure space and let $\mathcal{A}_1$ be the algebra which results from adjoining the function which is identically one on S to the algebra $\mathcal{A}$ of all real-valued bounded measurable functions on M, which vanish at infinity. It is easy to see from well-known representation theorems that $\mathcal{A}_1$ is algebraically isomorphic to the algebra $\mathcal{R}(W)$ of all real-valued continuous functions on a compact Hausdorff space W. Now either $\mathcal{A} = \mathcal{A}_1$ or $\mathcal{A}$ is a maximal ideal in $\mathcal{A}_1$ (according as M is finite or not), and hence A corresponds either to $\mathcal{R}(W)$, or to a maximal ideal in $\mathcal{R}(W)$. Now every such ideal consists of all elements of $\mathcal{R}(W)$ which vanish at some fixed point w_0 of W. Putting $S' = W$, or $S' =$ the space obtained from W by deletion of w_0, according as $\mathcal{A} = \mathcal{A}_1$ or not, it follows that $\mathcal{A}$ is algebraically isomorphic with the algebra $\mathcal{R}(S')$ of all real-valued continuous functions which vanish at infinity on the locally compact Hausdorff space S'.

Let ϕ be the algebraic isomorphism from $\mathcal{A}$ onto $\mathcal{R}(S')$. Putting $\mathcal{A}_0$ for the set of elements of $\mathcal{A}$ which vanish outside sets of finite measure and $\mathcal{R}_0(S')$ for the subalgebra of $\mathcal{R}(S')$ consisting of elements which vanish outside of compact sets, then $\phi(\mathcal{A}_0) = \mathcal{R}_0(S')$, for $\mathcal{A}_0$ and $\mathcal{R}_0(S')$ may be characterized purely algebraically as the collections of elements p in the respective algebras for which there is an element q in the algebra such that $pq = p$. We now define a functional L on $\mathcal{R}_0(S')$ by the equation $L(f) = \int_S \phi^{-1}(f)(x)\,ds(x)$. It is easy to verify that L is linear, and non-negative on non-negative functions (note that ϕ is order-preserving, for it maps squares into squares). Hence there exists a regular measure space $M' = (S', \mathcal{S}', s')$, where $\mathcal{S}'$ is contained in the σ-ring determined by the compact subsets of S', such that $L(f) = \int_{S'} f(x)\,ds'(x)$ (see e. g. [2]). We show next that M' is perfect.

To show that the characteristic function of any chunk in M' coincides n. e. with an element of $\mathcal{R}(S')$, let E' be a chunk in M', and put A for the linear functional on $\mathcal{A}$ defined by the equation $A(f) = \int_{E'} \phi(f)\,ds'(x)$. It is easy to see that A is linear, and that for $f \in A_0$, $|A(f)| \leq \|f\|_1$. Since $\mathcal{A}_0$ is dense in $L_1(M)$, it results that the contraction of A to $\mathcal{A}_0$ can be extended to a bounded linear functional A_1 on $L_1(M)$. By Theorem 5. 1 and the localizability of M there exists a bounded measurable function h on M such that $A_1(f) = \int_S f(x)h(x)\,ds(x)$, for $f \in L_1(M)$. It is plain that A is non-negative on non-negative functions, and hence that A_1 also is, from which it follows that h is n. e. non-negative-valued.

Now $h \in \mathcal{A}$, for let $F = [x \mid h(x) > \epsilon]$ where $\epsilon > 0$, and let $\{F_n; n = 1, 2, \cdots\}$ be an arbitrary monotone increasing sequence of chunks contained in F; to show that F is of finite measure, it suffices to show that the sequence $\{s(F_n)\}$ is bounded. Let f_n be the characteristic function of F_n, let $\{\alpha_n; n = 1, 2, \cdots\}$ be an arbitrary sequence of non-negative numbers for which the series $\sum_n \alpha_n$ is convergent, and set $f = \sum_n^{\infty} \alpha_n f_n$. Then it is easy to see that $f \in \mathcal{A}$, and for any positive integer n, $\sum_{1 \leq i \leq n} \alpha_i f_i \leq f$, which implies that $A(\sum_{1 \leq i \leq n} \alpha_i f_i) \leq A(f)$. It follows that $A(f) \geq \sum_{i=1}^{\infty} \alpha_i A(f_i) \geq \epsilon \sum_{i=1}^{\infty} \alpha_i s(F_i)$, so that the series $\sum_{i=1} \alpha_i s(F_i)$ is convergent. By virtue of the arbitrariness in the sequence $\{\alpha_i\}$, we can conclude that the sequence $\{s(F_i)\}$ is bounded. Hence, for all f in $\mathcal{A}_0$, $\int_{E'} \phi(f)(x')\, ds'(x) = \int_S f(x) h(x)\, ds(x) = \int_{S'} \phi(f)(x) \phi(h)(x)\, ds'(x)$, so that $\int_{S'} f'(x)(\chi_{E'}(x) - \phi(h)(x))\, ds'(x) = 0$ for all $f' \in \mathcal{R}_0(S')$. Since $\mathcal{R}_0(S')$ is dense in $L_1(M')$, and since the function $k'(x) = \chi_{E'}(x) - \phi(h)(x)$ is bounded, it follows that $\int_{S'} f'(x) k'(x)\, ds'(x) = 0$ for $f' \in L_1(M')$. Hence $k'(x) = 0$ n. e., i. e. $\chi_{E'}(x) = \phi(h)(x)$ n. e.

To show that M' is perfect, it remains only to show that if the chunk E' in M' has an interior point, then $s'(E') > 0$. Now if E' has an interior point, there exists a non-negative continuous function f' on S' which is one at the interior point and which vanishes outside E' and has values between zero and one. Then $\chi_{E'} \geq f'$, so that $L(\chi_{E'}) \geq L(f')$, i. e.

$$s'(E') \geq \int_S \phi^{-1}(f')(x)\, ds(x),$$

and since $f' \geq 0$ and $f' \neq 0$, we have $\phi^{-1}(f') \geq 0$ and $\phi^{-1}(f) \neq 0$, and hence $\int_S \phi^{-1}(f')(x)\, ds(x) > 0$.

That M is metrically equivalent to M' now follows from Theorem 4.3 combined with the following lemma (which is slightly more general than is necessary for the proof of the present theorem).

LEMMA 6.1.1. *For any bounded measurable function f on a perfect measure space there is a unique continuous function $\tilde{f}$ on the space which coincides n. e. with f. This function $\tilde{f}$ is necessarily bounded and vanishes at infinity as a function on a topological space if and only if f vanishes at infinity as a function on the measure space.*

It is readily seen to be sufficient to consider the case in which f is real-valued. Suppose to begin with that the perfect measure space $(T, \mathfrak{T}, t)$ on which f is defined is such that T is compact. Putting

$$E_{i,n} = [x \mid i2^{-n} > f(x) \geq (i-1)2^{-n}] \quad (i = 0, \pm 1, \pm 2, \cdots; n = 1, 2, \cdots),$$

then $E_{i,n}$ is a chunk, and there exists an element $f_{i,n}$ of $\mathfrak{R}(T)$ which is equivalent to the characteristic function of $E_{i,n}$. Setting $f_n = \sum_i i2^{-n} f_{i,n}$, where the summation is over all i for which $f_{i,n} \neq 0$, the boundedness of f implies that this summation is finite. It results that $f_n \in \mathfrak{R}(T)$ and that $|f_n(x) - f(x)| \leqq 2^{-n}$ n. e. for $x \in T$, and finally that $|f_n(x) - f_{n+1}(x)| \leqq 2^{-n+1}$ n. e. for $x \in T$.

Now the complement in T of a set of measure zero is dense in T, for otherwise there would exist a non-void open subset of T of zero measure. Hence, $f_n - f_{n+1}$ being a continuous function in T and the inequality $|f_n(x) - f_{n+1}(x)| \leqq 2^{-n+1}$ holding n. e., this inequality holds for all $x \in T$. It follows that the series $\sum_{n=1}^{\infty} \{f_{n+1}(x) - f_n(x)\}$ is uniformly convergent, i. e. the sequence $\{f_n\}$ converges uniformly, say to $\tilde{f}$, where $\tilde{f}$ is evidently in $\mathfrak{R}(T)$. Plainly $\tilde{f}(x) = f(x)$ n. e. on T. Moreover, $\tilde{f}$ is the only continuous function on T which is n. e. equal to f, for the difference of two such functions is n. e. zero, and so by the remark at the beginning of the paragraph is identically zero.

Now let $N = (T, \mathfrak{T}, t)$ be an arbitrary perfect measure space (so T need not be compact) and let f be an arbitrary bounded measurable function on N. If E is a chunk in N, there is a unique continuous function k on T which vanishes at infinity, and is equivalent to the characteristic function of E. It is plain that k^2 is equivalent to, and hence equal to, k. It results that k is the characteristic function of an open and closed subset E' of T, and since k vanishes at infinity on T, E' is compact. Now the contraction of f to E' is a bounded measurable function on the measure space $N_E = (E', \mathfrak{T}_E, t)$, where $\mathfrak{T}_E$ consists of the collection of all elements of T which are subsets of E. It follows from the preceding paragraph that there exists a continuous function f'_E on E', which is equivalent on N_E to the contraction of f (i. e. $f'_E(x) = f(x)$ n. e. for $x \in E'$).

Next we define $\tilde{f}$ on T as follows: if $x \in E'$, $\tilde{f}(x) = f'_E(x)$; $\tilde{f}$ is single-valued, for if $x \in E' \cdot F'$, where F is a chunk in N, then f'_E and f'_F are both defined and continuous on $E' \cdot F'$, and coincide n. e. on that set, and hence coincide everywhere there. It is clear that $\tilde{f}$ is continuous on T, and is

bounded on any E' by the essential bound of f, and so is bounded on T. It is obvious that $\tilde{f}$ is the only continuous function on T which coincides n. e. with f. If moreover $\tilde{f}$ vanishes at infinity as a function on a topological space, then $[x \mid |f(x)| \geq e]$ is equivalent to the compact set $[x \mid |f(x)| \geq e]$, for any positive number e, so that f vanishes at infinity as a function on a measure space. Finally, if conversely f vanishes at infinity on N, then it is not difficult to see that $\tilde{f}$ is the uniform limit of a sequence $\{f_n\}$ of finite linear combinations of characteristic functions of chunks. Now f_n is in $\mathfrak{R}(T)$, and $|f_n(x) - f_m(x)|$ is equivalent on N to $|f_n(x) - f_m(x)|$, so that l. u. b.$_{x \in T}|f_n(x) - f_m(x)| =$ Ess. l. u. b.$_{x \in T}|f_n(x) - f_m(x)|$. It results that the sequence $\{f_n\}$ converges uniformly. The limit is on the one hand equivalent to $\tilde{f}$ and on the other is a continuous function vanishing at infinity (being a uniform limit of continuous functions vanishing outside of compact sets). Thus f vanishes at infinity as a function on T.

It remains only to show that if two perfect spaces are metrically equivalent, then they are abstractly identical. Now if two measure spaces are metrically equivalent, the algebras of bounded measurable functions vanishing at infinity on them are algebraically isomorphic, by Theorem 4. 3. But in the case of a perfect space this algebra is by Lemma 6. 1. 1 naturally isometrically isomorphic to the algebra of all continuous functions on the space which vanish at infinity in the topological sense. Thus if $M = (R, \mathfrak{R}, r)$ and $M' = (R', \mathfrak{R}', r')$ are metrically equivalent perfect spaces, with the metric equivalence there is naturally associated an isometric isomorphism of the algebra of continuous functions vanishing topologically at infinity on R with the corresponding algebra on R'. By a theorem of Stone [7] (stated for compact spaces but extendable to locally compact spaces) this algebraic isomorphism can be implemented by a homomorphism of R with R', and there is no difficulty in showing that this homomorphism induces an isomorphism of M with M'.

The following theorem makes the results of the preceding section applicable to perfect spaces.

THEOREM 6. 2. *A perfect space is localizable.*

Let $N = (S, \mathfrak{S}, s)$ be a perfect space, and let $\mathfrak{K}$ be a conditional σ-ideal of chunks in N. By Theorem 5. 1, it suffices to show that $\mathfrak{K}$ is localized. Each chunk in $\mathfrak{K}$ is equivalent to an open and closed set in S, and the union of all such sets is an open set K. To show that $\mathfrak{K}$ is localized in K it is necessary only to prove that if E is a chunk which is contained in K, then $\bar{E} \in \bar{\mathfrak{K}}$. Since N is regular, there exists a sequence $\{C_n\}$ of compact subsets

of E such that $\bar{E} \subset \bigcup_n \bar{C}_n$. Now $C_n \subset K$, so that C_n is covered by the open and closed sets equivalent to the elements of K, and hence is covered by a finite number of such open and closed sets. It results that $\bar{C}_n \in \bar{\mathcal{K}}$, and hence $\bar{E} \in \bar{\mathcal{K}}$.

UNIVERSITY OF CHICAGO.

REFERENCES.

[1] J. Dieudonné, "Sur le theoréme de Lebesgue-Nikodym (III)," *Annales de l'Université de Grenoble*, vol. 23 (1947-48), pp. 24-56, esp. pp. 39-42.

[2] P. R. Halmos, *Measure theory* (New York, 1950).

[3] S. Kakutani, "Concrete representation of abstract (L) spaces and the mean ergodic theorem," *Annals of Mathematics*, vol. 42 (1941), pp. 523-537.

[3a] D. Maharam, "On homogeneous measure algebras," *Proceedings of the National Academy of Sciences*, vol. 28 (1942), pp. 108-111.

[4] I. E. Segal, "Decomposition of operator algebras. I." To appear in *Memoirs of the American Mathematical Society*.

[5] ———, "Decompositions of operator algebras. II: Multiplicity theory." To appear in *Memoirs of the American Mathematical Society*.

[6] M. H. Stone, "Notes on integration, IV," *Proceedings of the National Academy of Sciences*, vol. 35 (1949), pp. 50-58.

[7] ———, "Applications of the theory of Boolean rings to general topology," *Transactions of the American Mathematical Society*, vol. 41 (1937), pp. 375-481.

[8] J. von Neumann, "On rings of operators. III," *Annals of Mathematics*, vol. 41 (1940), pp. 94-161, esp. pp. 127-129.

[9] A. Weil, "Sur les espaces a structure uniforme," *Actualités scientifiques et industrielles*, No. 551 (Paris. 1937).

UNE EXTENSION DU THÉORÈME DE VITALI.*

Par ARNAUD DENJOY.

1. Dans la théorie de la mesure euclidienne des ensembles appartenant à un espace cartésien, le théorème publié par Vitali en 1907 [1] est parmi les plus importants. Ses applications aux propriétés des fonctions d'ensemble ou de variables réelles sont nombreuses. Par la promptitude, la simplicité des démonstrations utilisant ce théorème, celui-ci joue un role fondamental dans la théorie métrique des ensembles et des fonctions de nature cartésienne.

Il était naturel de chercher à étendre le théorème de Vitali, d'abord aux fonctions métriques quelconques pour les espaces cartésiens, ensuite aux métriques des espaces généraux.

Rappelons qu'une mesure euclidienne dans un espace cartésien U_r à r dimensions satisfait à cette condition que deux ensembles se déduisant l'un de l'autre par un déplacement euclidien, c'est-à-dire conservant les distances de deux points quelconques, ont même mesure, et que, dans le plan U_2, dans l'espace à trois dimensions U_3, etc., le carré, le cube, etc. de côté 1 ont l'aire, le volume, etc. égal à 1. Enfin, la mesure euclidienne utilisée par Vitali est une mesure borélienne, c'est-à-dire possédant l'additivité complète, et la soustractivité (voir au no. 6).

Généralement, une fonction d'ensemble, métrique $\phi(E)$, définie dans U_r est une quelconque fonction d'ensemble borélienne et non négative.

2. Toutes les tentatives d'élargir le champ de validité du théorème de Vitali ont sensiblement échoué. Nous allons en donner la raison.

En énonçant pour les espaces cartésiens, et plus précisément pour le plan U_2, les hypothèses de son théorème, c'est-à-dire des conditions suffisantes pour la validité de sa conclusion, Vitali a puisé trop largement dans les richesses de la nature métrique de ces espaces.

Dans l'espace U_r à r dimensions $(r > 2)$, pour lequel le théorème de Vitali se formule immédiatement comme son auteur l'avait énoncé pour le

* Received October 25, 1950.

[1] Atti Accad. Sci. Torino, 43, 75-92 (1908). Voir aussi S. Banach (*Fundamenta Math.*, 5, 130-136 (1924). La présente étude, exposée pour la première fois à l'Université Johns Hopkins en septembre 1950, a fait l'objet de trois notes aux *Comptes Rendus de l'Ac. des Sc. de Paris* (t. 251, 1950, pp. 560-562; 600-601; 737-739.

plan U_2, dans U_r *à la mesure euclidienne d'ordre r*, qui est l'aire pour $r = 2$, le volume pour $r = 3$, et *qui seule figure dans la conclusion du théorème*, s'ajoutent une mesure pour les *distances* entre deux points et pour les longueurs des ensembles, une mesure des aires, etc. Mais puisque la conclusion du théorème de Vitali ne fait état, pour U_r, que de la mesure euclidienne d'ordre r, on conçoit qu'en introduisant dans les hypothèses, en plus de cette première mesure, de surcroît la notion de distance entre les points, on augmente considérablement la difficulté d'étendre le théorème à des espaces métrisés non cartésiens ou à des espaces cartésiens à métrique non euclidienne. Car en général aux métriques spéciales il ne correspond par un lien de caractère simple et nécessaire aucune mesure des distances entre les points.

Ce n'est pas tout. Les ensemble spécifiés dans les hypothèses de Vitali non seulement remplissent deux conditions de nature métrique, quant à leurs volumes d'une part (pour $r = 3$), quant à leur dimension d'autre part (quelque soit r), mais il appartiennent encore à des espèces topologiques très particulières. Les uns sont des ensembles fermés, les autres ne diffèrent pas sensiblement d'ensembles ouverts. Or la conclusion du théorème de Vitali ne renferme aucune notion topologique.

3. Voici le théorème classique de Vitali pour l'espace à trois dimensions U_3:

THÉORÈME DE VITALI (énoncé A). *Si tout point M d'un ensemble H possédant un volume défini est centre d'une suite de parallélipipèdes rectangles* $\omega_n = \omega_n(M)$, *dont les côtés, tendant vers* 0 *quand n croît, ont leurs rapports mutuels compris entre deux nombres positifs k et* $1/k$ $(0 < k < 1)$ *indépendants de M et de n, et si chaque parallélipipède* ω [*identique à un* $\omega_n(M)$] *contient un ensemble* FERMÉ γ *dont le volume* $V(\gamma)$ *surpasse une fraction* $\theta V(\omega)$ *du volume de* ω, θ *étant indépendant de* ω, *il est possible de trouver parmi les* γ *une collection dénombrable d'ensembles* $\gamma_1, \gamma_2 \cdots, \gamma_i, \cdots$, DISJOINTS, *et dont la réunion* $\Gamma = \Sigma\gamma_i$ *couvre H, à un ensemble de volume nul prés. En outre on peut supposer que les points de* Γ *étrangers à H forment un ensemble de volume inférieur à un nombre positif indépendant* ϵ *préalablement donné.*

Etant d'abord bien spécifiée cette circonstance capitale que les γ_i sont *disjoints*, les conclusions du théorème tiennent en ces deux égalités, dont la première contient la partie essentielle de la proposition:

$$V(H - H \cdot \Gamma) = 0, \qquad V(\Gamma - H \cdot \Gamma) < \epsilon.^2$$

---

[2] Je garde les signes $(+, \Sigma)$ pour signifier l'addition des ensembles et $(\cdot, \Pi)$ pour

Dans cette double conclusion il n'est question que de *volumes* d'ensembles. Cette dernière notion s'étend immédiatement aux fonctions métriques dans U_3, fonctions donnant par exemple à tout ensemble ouvert borné O une mesure $\phi(O)$ finie et non toujours nulle. Aux volumes des ensembles de U_3 on peut faire pareillement correspondre dans des espaces très généraux U une fonction d'ensemble $\phi(E)$ borélienne non négative. Un théorème généralisant pour l'espace U celui de Vitali pour U_3, concernera deux familles [3] d'ensembles associés chacun à chacun (ω, γ), la famille P groupant les $\omega = \omega(\gamma)$ et la famille G réunissant les $\gamma = \gamma(\omega)$; ω donné contient $\gamma(\omega)$, γ donné est dans $\omega(\gamma)$. Le rapport $\phi(\gamma)/\phi(\omega)$ devra être borné inférieurement, indépendamment du couple associé (ω, γ). L'ensemble H sera l'ensemble des points M communs à une suite d'ensembles ω *tendant vers* 0 *en un certain sens à préciser*. On devra assujettir ces familles P, G à des hypothèses telles qu'en conclusion, il existe dans G une collection d'ensembles *disjoints* $\gamma_1, \gamma_2, \cdots$, dont la réunion $\Gamma = \Sigma\gamma_i$ donne l'égalité $\phi(H - H \cdot \Gamma) = 0$.

Quant à la condition, moins remarquable *a priori*, $\phi(\Gamma - H \cdot \Gamma) < \epsilon$, il sera également possible de l'obtenir.

L'essentiel est que nous chercherons à exprimer les hypothèses en termes n'invoquant point l'existence dans l'espace U d'une seconde nature métrique différente de celle que manifeste la fonction $\phi(E)$.

Enfin nos hypothèses n'exigeront pas plus que nos conclusions l'existence d'une topologie de l'espace U.

4. Réduisons d'abord dans la mesure du possible le caractère particulier des hypothèses du théorème de Vitali en son énoncé A.

Soit O un ensemble ouvert contenant H. Nous pouvons retrancher de la famille des parallélipipèdes ω tous ceux qui, frontière comprise, ne sont pas totalement intérieurs à O, sans que la condition caractérisant les points M de

---

leur intersection. Je n'écris $E_1 - E_2$ que si l'ensemble E_1 contient l'ensemble E_2. Sinon, je note $E_1 - E_1 \cdot E_2$ l'ensemble des points de E_1 étrangers à E_2. La somme de E_1 et de E_2 décomposée en ensembles disjoints est $E_1 + (E_2 - E_1 \cdot E_2)$ ou $E_2 + (E_1 - E_1 \cdot E_2)$. Une expression telle que $E = (E_1 - E_1 \cdot E_2)(E_3 - E_3 \cdot E_4)(E_5 - E_5 \cdot E_6) + \cdots$ se développe selon les règles du calcul algébrique. M étant un point de l'espace contenant les ensembles, on retient les ensembles $E_i \cdots E_j$ contenant M. La somme de leurs coefficients est toujours 1 ou 0, 1 si M est dans, E, 0 si M est étranger à E.

Les notations fréquemment utilisées aujourd'hui $\cup$, $\cap$ et $E_1 - E_2$, même si E_1 ne contient pas E_2, ne se prêtent pas à l'application de règles aussi simples et précises.

[3] Nous distinguons d'une part les *familles* P, G d'ensembles ω, γ, P et G étant respectivement l'ensemble dont les ω, les γ constituent les éléments, d'autre part les *sommes* ou *réunions* de collections d'ensembles ω, γ. Celles-ci notées $\Sigma\omega$, $\Sigma\gamma$ sont les ensembles formés par les éléments appartenant aux divers ensembles ω, γ respectivement réunis.

H cesse d'être vérifiée par aucun de ces points. En effet, O étant ouvert, M situé dans O est à une distance positive 2δ de la frontière de O. A partir d'une certaine valeur de n, les côtés du parallélipipède $\omega_n(M)$ sont inférieurs à δ. Ces parallélipipèdes et leurs frontières sont dès lors intérieurs à O. Ils n'ont pas été compris dans les suppressions effectuées parmi les ω et qui, pour chaque point M ne font disparaitre qu'un nombre fini de parallélipipèdes $\omega_n(M)$.

Si H a un volume fini, et même si le volume de H est simplement déterminé à toute distance finie (mais infiniment grand avec cette distance), H peut être compris dans un ensemble ouvert O tel que $V(O-H) < \epsilon$. Dès lors, si l'on réduit les ω à ceux qui, avec leurs frontières, sont inclus dans O, tous les γ correspondants sont *a fortiori* inclus dans O, et il en est de même de tout ensemble ponctuel $\Gamma = \Sigma\gamma_i$ réunissant une collection d'ensembles γ. Mais Γ étant dans O, l'ensemble $\Gamma - \Gamma \cdot H$ des points de Γ étrangers à H est dans l'ensemble $O - H$ des points de O étrangers à H. Donc $V(\Gamma - \Gamma \cdot H) < \epsilon$.

Il suit de là que la démonstration de l'énoncé A du théorème de Vitali se réduit à celle du premier point, l'existence de la collection d'ensembles disjoints γ_i tels que, avec $\Gamma = \Sigma\gamma_i$, $V(H - H \cdot \Gamma) = 0$.

Enfin, si la démonstration est acquise pour le cas où $V(H)$ *est fini*, la conclusion vaudra aussi quand le volume de H, déterminé à toute distance finie (dans toute sphère ayant son centre à l'origine) ce volume devient infini avec cette distance. d désignant la distance d'un point de l'espace U_3 à l'origine, soit S_n la région intersphérique ouverte $(n-1 < d < n)$ et H_n l'ensemble $H \cdot S_n$, n prenant toutes les valeurs entières positives. $H - \Sigma H_n$ est formé de points situés sur les surfaces sphériques $d = n$, dont le volume total est nul. Donc $V(H - \Sigma H_n) = 0$. Soit ϵ_n un nombre positif terme d'une série dont la somme est inférieure à ϵ. Chacun des H_n ayant un volume fini, par hypothèse [la validité de l'énoncé A étant admise pour le cas de $V(H)$ fini], il existe pour chacun des H_n une famille $\Gamma_n = \sum_i \gamma_i^n$ d'ensembles $\gamma_1^n, \gamma_2^n, \cdots$ disjoints et inclus dans S_n, vérifiant $V(H_n - H_n \cdot \Gamma_n) = 0$, et même en outre $V(\Gamma_n - H \cdot \Gamma_n) < \epsilon_n$; Γ_n et Γ_m sont disjoints si $m \neq n$. La collection $\Gamma = \Sigma\Gamma_n = \sum_{i,n} \gamma_i^n$ est formée d'ensembles γ disjoints, et $V(H - H \cdot \Gamma) = 0$, $V(\Gamma - H \cdot \Gamma) < \epsilon$.

En résumé, le théorème de Vitali sous son énoncé A ne perd ni en généralité quand on suppose $V(H)$ fini, ni en portée si l'on néglige sa seconde conclusion $V(H - H \cdot \Gamma) < \epsilon$.

A ce théorème ainsi restreint et que nous appelons le *théorème A'*, nous allons substituer un autre énoncé B, renfermant moins de conditions de détail et parfaitement équivalent au premier.

5. Nous dirons qu'une famille P' d'ensembles ω' situés dans l'espace U *couvre indéfiniment au sens de la dimension* un point M, s'il existe une suite d'ensembles $\omega_1'(M), \omega'_2(M), \cdots$ appartenant à P', contenant M et dont les diamètres tendent vers 0.

THÉORÈME DE VITALI (énoncé B). *P' désignant une famille de sphères ouvertes ω' dont chacune contient un ensemble fermé $\gamma' = \gamma'(\omega')$ tel que $V(\gamma') > \theta' V(\omega')$ si $\gamma' = \gamma'(\omega')$, et le nombre positif θ' étant indépendant de ω', si l'ensemble H' indéfiniment couvert au sens de la dimension par la famille P' a un volume $V(H')$ déterminé, positif et fini, H' peut, à un ensemble de volume nul près, être couvert par une collection dénombrable $\gamma'_1, \gamma'_2, \cdots, \gamma'_i, \cdots$ d'ensembles γ'.*

Nous allons successivement supposer démontré le théorème sous sa forme A', puis sous sa forme B, et nous prouverons que la validité de la forme justifiée entraîne la validité de l'autre forme.

1° L'énoncé A' est supposé exact. Démontrons que B l'est aussi. Nous adoptons les hypothèses du théorème B. Englobons d'abord H' dans un ensemble ouvert O' de volume fini et dépassant si l'on veut de moins de ϵ le volume de H'. Nous ne changeons pas l'ensemble H' indéfiniment couvert par les ω' en supprimant d'entre ceux-ci toute sphère ω' telle que la sphère concentrique et de rayon 4 fois plus grand ne soit pas en totalité dans O'. La famille P' est réduite à une famille P'_0. Cela fait, si M' est un point de H' et si ω' est une sphère contenant M' et appartenant à P'_0, soit ω le cube de centre M' et de côtés (dirigés, si l'on veut, parallèlement aux axes de coordonnées) égaux au quadruple du rayon de ω'; ω contient ω'; ω est intérieur à O'. Au cube ω associons l'ensemble fermé $\gamma = \gamma(\omega)$ identique à $\gamma'(\omega')$; $\gamma(\omega)$ est dans ω, puisque $\gamma(\omega)$ est dans ω'. Le rapport $V(\gamma)/V(\omega)$ surpasse θ tel que θ/θ' soit égal au rapport de $V(\omega')$ à $V(\omega)$, soit $4\pi/3.64$; θ est indépendant de ω. Donc en vertu de l'énoncé A', H peut, à un ensemble de volume nul près, être couvert par une collection dénombrable de γ_i disjoints. Or ces γ_i sont des γ'_i de G'. L'énoncé B est donc une conséquence de l'énoncé A'.

2° L'énoncé B est supposé démontré. Reprenons les hypothèses de l'énoncé A' et montrons l'exactitude de sa conclusion. Si un point M de H est centre d'un parallélipipède rectangle ω de P, le plus grand côté l de celui-ci est tel que $l^3k^2 < V(\omega) < l^3$. Soit ω' la sphère de centre M et de rayon égal à $2l$; ω' contient ω. Nous pouvons avoir inclus préalablement H dans un ensemble ouvert 0 et avoir exclu de la famille des parallélipipèdes ω ceux dont le plus grand côté surpasse la moitié de la distance de M à la frontière de 0; ω' est dans O. A ω' nous faisons correspondre l'ensemble fermé $\gamma' = \gamma'(\omega')$

identique à $\gamma(\omega)$; γ' est dans ω' et $V(\gamma')/V(\omega') > \theta'$, si θ'/θ est le minimum du rapport $V(\omega)/V(\omega')$, soit $3k^2/32\pi$; θ' est indépendant de ω'. Soit H' l'ensemble indéfiniment couvert, au sens de la dimension, par la famille des ω'; H' contient H (et d'ailleurs H' est dans O dont le volume peut être supposé supérieur à celui de H de moins de ϵ). En vertu du théorème B, nous pouvons couvir H', sauf éventuellement un ensemble de volume nul, par une collection d'ensembles disjoints $\gamma'_i = \gamma_i$. A fortiori H est-il couvert par les γ_i à un ensemble de volume nul près. L'énoncé A' est justifié.

En résumé l'énoncé B du théorème de Vitali équivaut entièrement à l'énoncé A. L'énoncé B réduit le théorème de Vitali à ses hypothèses essentielles, et c'est sous cette forme B que nous entendrons ci-après le théorème originel. Dans l'énoncé B donné plus haut nous prions le lecteur de supprimer les accents des lettres P', ω', γ', H', qui deviendront comme dans les énoncés A et A', et dans l'énoncé général ultérieur: P, ω, γ, H.

6. Dans le Théorème Général que nous voulons substituer au théorème de Vitali pour le rendre applicable à des espaces variés U, nos hypothèses ne feront appel ni à d'autre notion métrique pour U que celle de l'existence de la fonction $\phi(E)$, ni à une topologie de U. Dans certains cas, celui des espaces cartésiens par exemple, la fonction $\phi(E)$ se définit à partir de ses valeurs pour une famille d'ensembles élémentaires [dans U_3, ce sont les cubes précisés C_k, savoir: $p_i \leq 2^k x_i < p_i + 1$, $i = 1, 2, 3$; p_i entier quelconque]. La définition de $\phi(E)$ s'étend à une classe majeure d'ensembles qualifiés de mesurables-ϕ, par progrès borélien, c'est-à-dire par application de ces deux principes: Si les ensembles E_n sont disjoints et si tous les $\phi(E_n)$ existent, pour l'ensemble $E = \Sigma E_n$ le nombre $\phi(E)$ existe et vaut $\Sigma\phi(E_n)$. D'autre part, si $\phi(E)$ et $\phi(E')$ existent et si $E \supset E'$, $\phi(E - E')$ existe, et des lors vaut $\phi(E) - \phi(E')$. En conséquence si, le nombre $\phi(E_n)$ existant quel que soit n, l'ensemble E_n est soit constamment croissant, soit constamment décroissant, et si E est sa limite (au sens abstrait: $E = \Sigma E_n$ dans le premier cas, $E = \Pi E_n$ dans le second) $\phi(E)$ existe et vaut $\lim \phi(E_n)$, [toutefois quand E_n est décroissant, il faut supposer $\phi(E_n)$ fini, du moins à partir d'une certaine valeur de n].

La soustractivité d'une mesure borélienne s'entend ainsi au sens général. Quels que soient les ensembles E et E' non disjoints et tels que $\phi(E)$, $\phi(E')$ existent:

1° $\phi(E - E \cdot E')$ existe. En conséquence, d'après

$$E \cdot E' = E - (E - E \cdot E'),$$

$\phi(E \cdot E')$ existe, et aussi, de par $E + E' = E + (E' - E \cdot E')$, $\phi(E + E')$ existe;

$$2^\circ \quad \phi(E - E \cdot E') = \phi(E) - \phi(E \cdot E').$$

Enfin, quel que soit l'ensemble E, il existe une mesure extérieure-ϕ notée $\phi_e(E)$ et qui est la borne inférieure stricte des mesures $\phi(E')$ des ensembles E' mesurables-ϕ et contenant E. La mesure intérieure $\phi_i(E)$ se définit de même par la borne supérieure stricte des mesures $\phi(E'')$ des ensembles E'' mesurables-ϕ intérieurs à E.

Les mesures extérieures et intérieurs $\phi_e(E)$ et $\phi_i(E)$ verifient une sorte d'additivité complète, se rattachant immédiatement à l'additivité complète des mesures-ϕ exactes. E étant la somme des ensembles E_n, si E_n est inclus dans un ensemble K_n mesurable-ϕ, et si les K_n sont disjoints, il en résulte $\phi_e(E) = \Sigma\phi_e(E_n)$; $\phi_i(E) = \Sigma\phi_i(E_n)$.

7. Pour aider le lecteur à pénêtrer pleinement le sens des hypothèses qui fonderont le Théorème Général, nous analysons ces conditions par avance.

Étant donnée une famille P d'ensembles ω non nécessairement mesurables-ϕ, nous dirons qu'*un point M de U est indéfiniment couvert, au sens de la métrique-ϕ,* par la famille P, s'il existe une suite d'ensembles ω_n contenant M et *dont la mesure extérieure* $\phi_e(\omega_n)$ *tend vers* 0 quand n croît.

Si dans U_3, $\phi(E)$ est le volume $V(E)$, un ensemble indéfiniment couvert par une famille d'ensembles ω au sens de la dimension l'est aussi au sens des volumes. Mais la réciproque n'est nullement vraie; $V(\omega)$ peut tendre vers 0 sans qu'il en soit ainsi du diamètre de ω. L'ensemble d'accumulation d'une suite d'ensembles ω_n, même ouverts, ayant un point commun et dont le volume tend vers 0 peut être identique à la totalité de l'espace U_3.

Donnons pour le plan U_2 et les aires $A(E)$ un exemple aisé à étendre à l'espace U_3 et aux volumes $V(E)$.

La région $u(\epsilon)$ comprise entre les courbes $y = -\epsilon/(1+x^2)$ et $y = \epsilon/(1+x^2)$ est finie et son aire est inférieure à 4ϵ. Cette région est coupée par la circonférence c ayant son centre à l'origine et le rayon 1 suivant un arc $a(\epsilon)$ sensiblement de longueur ϵ, d'ordonnées extrêmes $-\frac{1}{2}\epsilon$ et $\frac{1}{2}\epsilon$, d'abscisses voisines de 1. Soit $u'(\epsilon)$ la partie de $u(\epsilon)$ non intérieure à c. Sur c plaçons un ensemble parfait totalement discontinu dont les arcs contigus $a_1, a_2, \cdots, a_i, \cdots$ ont pour longueurs $\epsilon_1, \epsilon_2, \cdots, \epsilon_i, \cdots$; $\Sigma\epsilon_i \leq 2\pi$. En faisant tourner autour de l'origine $u'(\epsilon_i)$ ou plutôt $u'(\epsilon'_i)$, ϵ'_i différant peu de ϵ_i, nous appliquons l'arc $a(\epsilon'_i)$ sur l'arc a_i. A l'ensemble de régions ainsi obtenu ajoutons l'intérieur de c et les arcs a_i. Nous obtenons une région ω_1

d'aire inférieure à 9π; ω_0 évoque l'idée d'un poulpe dont l'intérieur de c formerait le corps et dont les régions infinies greffées sur les a_i seraient les tentacules, infinies en nombre et en dimension.

Dans ω_0 nous plaçons un ensemble fermé γ_0 obtenu en réduisant les rayons vecteurs r des points de ω_0 à $r' = r/(1+\alpha r)$ $(0 < \alpha < 1$, α étant indépendant de $r)$, et en ajoutant à cette région sa frontière. $A(\gamma_0)/A(\omega_0)$ tend vers 1 quand α tend vers 0.

Si nous prenons les homothétiques de ω_0 et de γ_0 avec un rapport λ nous obtenons des ensembles $\omega_0(\lambda)$, $\gamma_0(\lambda)$. Soit $z_n = x_n + iy_n$ un ensemble partout dense (par exemple x_n et y_n prennent indépendamment l'un de l'autre toutes les valeurs réelles rationnelles). Soit λ_n tels que $\Sigma\lambda_n^2$ converge et soit inférieur à 1. Si θ est incommensurable avec π, imprimons dans U_2 figurant le plan complexe, à $\omega_0(\lambda_n)$ et à $\gamma_0(\lambda_n)$ la translation z_n et la rotation $n\theta$. La réunion de tous ces ensembles ouverts respectivement égaux aux $\omega_0(\lambda_n)$ est un ensemble ouvert ω_1 d'aire inférieure à 9π et la réunion des ensembles égaux aux $\gamma_0(\lambda_n)$ est un ensemble fermé à distance finie, que nous arrêtons à une distance assez grande de l'origine pour que l'ensemble fermé γ_1 restant, et qui est inclus dans ω_1, ait une aire supérieure à $\frac{1}{2}A(\omega_1)$.

Le complémentaire de ω_1 est un ensemble, fermé à distance finie, mais vraisemblablement totalement discontinu, bien que ω_1 ait une aire finie, inférieure à 9π. Des homothéties appliquées à ω_1 permettent de réduire cette aire au-dessous de toute limite. On conçoit qu'en constituant une famille P avec des ensembles ω se rattachant à ω_1 et à ses homothétiques, par reproduction complète ou par découpage arbitraire, on conçoit qu'avec ces ensembles ω dont chacun peut être partout dense dans le plan U_2, la notion d'ensemble indéfiniment couvert par ces ω, au sens des aires, s'éloigne de celle de l'ensemble indéfiniment recouvert au sens de la dimension.

Nous avons cherché à compliquer la tâche d'imaginer l'ensemble H indéfiniment couvert au sens des aires par une famille P d'ensembles ω. Mais il est facile d'obtenir un ensemble ouvert ω d'aire finie, arbitrairement petite, le complémentaire de ω étant totalement discontinu, ce qui implique évidemment la densité de ω dans tout le plan. Les deux familles de droites parallèles aux axes Ox, Oy et dont la coordonnée constante est rationnelle forment un ensemble d'aire nulle. On peut donc inclure celui-ci dans un ensemble ouvert ω d'aire inférieure à ϵ. Le complémentaire de ω est évidemment totalement discontinu.

8. Revenons à nos hypothèses sur les familles ω, γ. Nous avons dit que l'ensemble H et les ensembles ω désignés dans l'un ou l'autre des énoncés

A et B du théorème de Vitali remplissent, au sens des volumes jouant le rôle de fonction métrique-ϕ, la condition de recouvrement indéfini.

Seulement dans ce théorème de Vitali, les ω qui sont des parallélipipèdes, comprenant ou non leurs frontières, et les γ qui sont des ensembles fermés, sont à la fois euclidiennement mesurables en volumes, et en outre doués de caractères topologiques remarquables.

Les ensembles ω de notre énoncé ne seront pas nécessairement mesurables-ϕ, leur mesure extérieure intervenant seule.[4] Les γ au contraire *seront nécessairement mesurables-ϕ, et de mesure positive* mais ni les ω ni les γ n'ont à remplir d'autre condition de nature, l'*espace U n'étant pas obligatoirement doué d'une topologie.*

9. Voici les conditions que nous allons imposer aux systèmes (ω, γ) pour assurer la validité du théorème général que nous avons en vue.

1° Quel que soit l'ensemble γ de la famille G, les points étrangers à γ et indéfiniment couverts au sens de la métrique-ϕ par les ω' joints à γ forment un ensemble $\rho(\gamma)$ de mesure-ϕ égale à zéro.

Supposons choisi l'ensemble γ_1 désigné dans la conclusion du théorème à énoncer. Il s'agira de trouver les autres γ_i $(i \geq 2)$. Avec la simple hypothèse $\gamma_i \subset \omega(\gamma_i)$, nous ne sommes assurés que les γ_i sont disjoints de γ_1 que si les $\omega(\gamma_i)$ sont eux-mêmes disjoints de γ_1. Donc ces $\omega(\gamma_i)$ seront dans la famille P_1 des ensembles ω restant de la famille P quand on retranche de celle-ci les ω joints à γ_1. *A priori* ces ω joints à γ_1 peuvent couvrir indéfiniment outre les points de $H \cdot \gamma_1$ ceux d'un ensemble $\rho(\gamma_1)$ appartenant à H. Il est possible que $\rho(\gamma_1)$ soit totalement disjoint des ω formant P_1 et dans ce cas la collection Γ des ensembles γ_i $(i \geq 1)$ ne contiendrait aucun point de $\rho(\gamma_1)$. La conclusion du théorème ne peut donc pas être générale si $\rho(\gamma_1)$ a une mesure extérieure-ϕ positive. Et c'est ainsi que la nécessité de notre première condition apparaît.

Evidemment $\rho(\gamma)$, s'il existe, est dans H, puisque tout point de $\rho(\gamma)$ est indéfiniment couvert (au sens de la métrique-ϕ) par une famille partielle ω' incluse dans P.

Avec les hypothèses de l'énoncé B de Vitali, γ est un ensemble fermé con-

---

[4] Quand l'espace U est cartésien, la considération d'ensembles ω non mesurables (nonboréliens) n'est pas d'une très grande utilité. Une des idées les plus importantes dues à Lebesgue, est que pratiquement l'Analyse ne peut conduire qu'à des ensembles mesurables, puisque les passages à la limite conservent ce caractère. On peut toutefois songer à des sommes non dénombrables d'ensembles mesurables, sommes qui eventuellement ne seraient pas mesurables.

tenu dans une sphère $\omega(\gamma)$. Les sphères ω' jointes à γ et dont le volume tend vers 0 ont évidemment leurs points d'accumulation dans γ. Elles ne peuvent donc couvrir indéfiniment aucun point étranger à γ; $\rho(\gamma)$ n'existe pas dans l'hypothèse de Vitali. La condition $V[\rho(\gamma)] = 0$ est verifiée.

Mais si, dans U_3 et avec la métrique des volumes euclidiens, chacun des ω peut être partout dense, on conçoit que cette première hypothèse introduit une restriction effective. La démonstration en fera connaître le rôle essentiel.

Soit M un point de U étranger à $\gamma + \rho(\gamma)$. Les ω' contenant M et joints à γ ont leur mesure extérieure-ϕ bornée inférieurement par un nombre β. Il n'y a pas d'ensemble ω' joint à γ et contenant M, et pour lequel $\phi_e(\omega') < \beta$. Si M est en même temps dans H, il y une infinité d'ensembles ω'' contenant M et tels que $\phi(\omega'')$ tende vers 0. Mais ces ω'' sont tous disjoints de γ dès que $\phi(\omega'') < \beta$.

2° Quel que soit γ dans G, il existe deux nombres a et b, vérifiant $1 < a < b$, indépendants de γ, et tels que les $\omega'(\gamma')$ joints à γ et vérifiant la condition $\phi(\gamma') < a\phi(\gamma)$ forment par leur réunion un ensemble $\Omega(\gamma)$ dont la mesure extérieure-ϕ est inférieure à $b\phi(\gamma)$.

$\omega(\gamma)$ est évidemment un des ω' puisque $\omega(\gamma)$ est joint à γ qu'il contient et que $\phi(\gamma) < a\phi(\gamma)$, d'après $a > 1$. Donc $\omega(\gamma) \subset \Omega(\gamma)$ et $\phi_e[\omega(\gamma)] < b\phi(\gamma)$.

En conséquence, si M est étranger à $\gamma + \rho(\gamma)$, les γ' tels que $\omega(\gamma')$ soit joint à γ et contienne M ont leur mesure $\phi(\gamma')$ supérieure à un nombre positif α, dépendant évidemment de M et de γ. Car d'une inégalité $\phi_e[\omega(\gamma')] > \beta$ résulte $\phi(\gamma') > \beta/b$, ce dernier nombre pouvant tenir le rôle de α.

Sous la forme B du théorème de Vitali, il est aisé de voir que la seconde de nos hypothèses est vérifiée quel que soit $a > 1$. De plus, tandis que la connaissance des nombres a et b, de $\phi(\gamma)$ et d'un point N de γ ne suffit pas à définir dans U un ensemble contenant certainement $\Omega(\gamma)$ indépendamment de la donnée de la famille P des ω, puisque ceux-ci peuvent être partout denses dans un espace U, au contraire dans les conditions du théorème B de Vitali, la connaissance de a, du coefficient θ, du volume $V(\gamma)$ et d'un point de γ permet de définir une sphère ayant ce dernier point pour centre et contenant $\Omega(\gamma)$, indépendamment de la détermination particulière de la famille P des sphères ω considérées dans l'énoncé B de Vitali.

En effet, d'après l'hypothèse $V(\gamma) > \theta V(\omega)$, ω étant $\omega(\gamma)$, la sphère ω a un rayon inférieur à $r = \{3V(\gamma)/(4\pi\theta)\}^{1/3}$. Si le point N est dans γ, la totalité de γ est dans la sphère de centre N et de rayon $2r$. Le nombre a étant quelconque supérieur à 1, si $V(\gamma') < aV(\gamma)$, $\omega(\gamma')$ a un rayon inférieur à $r' = a^{1/3}r$, et si en outre la sphère $\omega' = \omega(\gamma')$ est jointe à γ, elle est entière-

ment comprise dans la sphère de centre N et de rayon $2(1+a^{1/3})r$. Cette sphère contient l'ensemble $\Omega(\gamma)=\Sigma\omega'$. Son volume est inférieur au produit de $V(\gamma)$ par un nombre b fonction de a et de θ, indépendant de la composition des familles P et G. Cette sphère contenant $\Omega(\gamma)$ est déterminée par la connaissance des seuls nombres a, θ, $V(\gamma)$ et du point N de γ, centre de cette sphère.

3° La mesure extérieure-ϕ de l'ensemble D réunissant les ensembles γ est finie.

Nous montrerons plus loin (16) la nécessité de cette troisième hypothèse pour la validité du théorème général.

Dans le cas due théorème de Vitali sous sa forme B, l'ensemble indéfiniment couvert H a un volume fini et tout ensemble ouvert O contenant H, O ayant par exemple un volume excédant celui de H de moins de ϵ, peut être considéré comme contenant tous les ensembles γ.

Les trois hypothèses que nous retenons sont donc satisfaites dans le cas particulier régi par le théorème de Vitali.

10. Voici maintenant notre énoncé:

THÉORÈME GÉNÉRAL. *Dans l'espace U où une fonction d'ensemble $\phi(E)$ non négative, borélienne, est définie et sert de mesure, on suppose données deux familles d'ensembles (ω,γ) associés, P famille d'ensembles $\omega=\omega(\gamma)$ non nécessairement mesurables-ϕ, G famille d'ensembles $\gamma=\gamma(\omega)$, $\gamma(\omega)$ étant inclus dans ω et possédant une mesure $\phi(\gamma)$ déterminée, positive et finie,* et ces familles vérifiant les conditoins suivantes:

1° *Quel que soit l'ensemble γ dans G, les points étrangers à γ et indéfiniment couverts au sens de la métrique-ϕ par les ω' de P joints à γ forment un ensemble $\rho(\gamma)$ dont la mesure-ϕ est nulle;*

2° *il existe deux nombres positifs a et b $(1<a<b)$ indépendants de γ quelconque dans G et tels que l'ensemble $\Omega(\gamma)$ réunissant les ω' joints à γ et vérifiant $\phi(\gamma')<a\phi(\gamma)$ si $\gamma'=\gamma'(\omega')$ a une mesure extérieure-ϕ inférieure à $b\phi(\gamma)$.*

3° *La mesure extérieure-ϕ de l'ensemble D réunissant les γ est finie;*

sous ces conditions suffisantes, *H désignant l'ensemble des point M de U indéfiniment couverts au sens de la métrique-ϕ par la famille P des ensembles ω* [M appartient à une suite d'ensembles ω_n de P dont la mesure extérieure $\phi_e(\omega_n)$ tend vers 0], *et Δ, inclus dans H, étant l'ensemble des points indéfiniment couverts au sens de la métrique-ϕ par la famille G des γ:*

1° *L'ensemble H est couvert, à un ensemble de mesure-ϕ nulle près,*

par une collection dénombrable d'ensembles $\gamma_1, \gamma_2, \cdots$ *de* G, *disjoints. Si* $\Gamma = \Sigma\gamma_i$, $\phi(H - H \cdot \Gamma) = 0$;

2° H *et* Δ *sont mesurables-*ϕ, *l'excès de* H *sur* Δ *ayant la mesure-*ϕ *nulle* $[\phi(H) = \phi(\Delta)]$. *De plus, si petit que soit le nombre positif* ϵ, *on peut choisir les* γ_i *de* G *de façon que* $\phi(\Delta) \leq \phi(\Gamma) < \phi(\Delta) + \epsilon$.

11. Pour démontrer la première conclusion de ce théorème, nous suivrons la marche indiquée par Banach pour établir le théorème classique de Vitali.

De $D = \Sigma\gamma$ résulte $\phi(\gamma) < \phi_e(D)$, quel que soit γ dans G; d'après la troisième hypothèse: $[\phi_e(D) < \infty]$, les nombres $\phi(\gamma)$ sont bornés supérieurement. Soit μ_1 leur maximum (borne supérieure stricte) et γ_1 un ensemble γ de G tel que $\phi(\gamma_1) > \mu_1/a$; γ_1 existe d'après $a > 1$. Quel que soit γ dans G, $\phi(\gamma) \leq \mu_1 < a\phi(\gamma_1)$.

Il peut se faire que précisément les points de H étrangers à γ_1 forment un ensemble de mesure-ϕ nulle. Si $\phi(H - H \cdot \gamma_1) = 0$, la proposition est vérifiée. Supposons qu'il en soit autrement.

Posons $\rho_1 = \rho(\gamma_1)$; ρ_1 est dans H, et $H \cdot \rho_1 = \rho_1$; d'après la première hypothèse, $\phi(\rho_1) = 0$. Soit $H_1 = H - H(\gamma_1 + \rho_1)$, ou $H_1 = H - H \cdot \gamma_1 - \rho_1$; H_1 et $H - H \cdot \gamma_1$ ont la même mesure extérieure, et nous supposons que celle-ci est positive. Si M est étranger à $\rho_1 + \gamma_1$, il existe un nombre α_1, dépendant de M (et de γ_1 évidemment) tel que si $\phi(\gamma) < \alpha_1$ et si $\omega(\gamma)$ contient M, $\omega(\gamma)$ n'est pas joint à γ_1. Soit M un point de H_1; H_1 étant dans H, M appartient à H, donc M est contenu dans une suite d'ensembles ω de mesure extérieure tendant vers 0. Et M inclus dans H_1 est étranger à $\gamma_1 + \rho_1$. Donc dès que, ω contenant M, $\phi_e(\omega)$ est assez petit pour que $\phi(\gamma) < \alpha_1$, γ étant $\gamma(\omega)$, toute la suite restante des ω est disjointe de γ.

De P retranchons tous les ω joints à γ_1 et soit P_1 la famille restante. Nous venons de montrer qu'au sens de la métrique-ϕ, la famille P_1 couvre indéfiniment tout point de H_1. Soit H'_1 l'ensemble indéfiniment couvert, au sens de la métrique-ϕ, par la famille P_1; H'_1 contient H_1, nous venons de le dire. D'ailleurs H'_1 est évidemment inclus dans H, de même que la famille P_1 est incluse dans la famille P; tout point indéfiniment couvert par P_1 l'est *a fortiori* par P; Mais comme P_1 ne contient aucun ω joint à γ_1, H'_1 ne renferme aucun point de γ_1; H'_1 inclus dans H est disjoint de $H \cdot \gamma_1$; H'_1 contenant $H_1 = H - H \cdot \gamma - \rho_1$ est contenu dans $H - H \cdot \gamma_1 = H_1 + \rho_1$. ($H_1$ est disjoint de ρ_1, puisque H_1 se déduit de H par extraction non seulement des points appartenant à γ_1, mais encore de tous les points de ρ_1, ceux-ci étant étrangers à γ_1). Les points de H'_1 étrangers à H_1 sont donc dans ρ_1. Si

$H'_1 = H_1 + \rho'_1$, ρ'_1, comme ρ_1 dont il est une partie, a la mesure-ϕ nulle. D'ailleurs ρ'_1 peut exister, H'_1 peut contenir des points de ρ_1. Il est évidemment possible qu'un point étranger à γ_1 et couvert indéfiniment (au sens de la métrique-ϕ) par les ω de P joints à γ_1 soit aussi couvert indéfiniment par les ω de P disjoints de γ_1, et qui forment P_1.

Soit G_1 l'ensemble des $\gamma = \gamma(\omega)$ correspondant aux ω de P_1. Puisque tout γ de G_1 est dans un ω disjoint de γ_1, *a fortiori* tous les γ de G_1 sont disjoints de γ_1. D'autre part le maximum μ_2 de $\phi(\gamma)$ pour les γ de G_1 vaut au plus μ_1 et il est inférieur à $a\phi(\gamma_1)$.

12. Maintenant, avec les familles P_1 et G_1 nous procédons comme nous l'avons fait avec P et G. Cette règle engendre une suite d'opérations qui nous donnent une succession d'ensembles $\gamma_1, \gamma_2, \cdots, \gamma_i, \cdots$ disjoints, se déterminant (même à partir de γ_1, si $P = P_0$, $G = G_0$) de la façon suivante:

Soit $\Gamma_n = \gamma_1 + \gamma_2 + \cdots + \gamma_n$. Si $H - H \cdot \Gamma_n$ a la mesure-ϕ nulle, la proposition est vérifiée. Supposons que cet ensemble ait une mesure extérieure-ϕ positive. Posons $\rho_i = \rho(\gamma_i)$. D'après la première hypothèse, l'ensemble $\rho_1 + \rho_2 + \cdots + \rho_n = \Theta_n$ a la mesure-ϕ nulle. Il est inclus dans H. Mais Θ_n n'est pas nécessairement disjoint de Γ_n. Car ρ_i, disjoint de γ_i, ne l'est pas nécessairement de tous les γ_j pour $j \neq i$. Soit $H_n = H - H \cdot (\Gamma_n + \Theta_n)$. Par hypothèse $\phi_e(H_n) > 0$.

Soit M un point quelconque de H_n; M étant dans H, il existe une suite d'ensembles ω contenant M et tels que $\phi_e(\omega)$ tende vers 0. D'autre part, M étant étranger à $\Gamma_n + \Theta_n$ est étranger à tout ensemble $\gamma_i + \rho_i$ pour $1 \leq i \leq n$. Donc il existe un nombre α_i dépendant de M et de γ_i, tel que si $\phi(\gamma) < \alpha_i$ et si $\omega = \omega(\gamma)$ contient M, $\omega(\gamma)$ est disjoint de γ_i.

Soit α'_n est le plus petit des nombres $\alpha_1, \alpha_2, \cdots, \alpha_n$. Considérons les ω vérifiant ces conditions: ω contient M et $\phi(\gamma) < \alpha'_n$ si $\gamma = \gamma(\omega)$. Ces ω sont tous disjoints de Γ_n et ils renferment parmi eux une suite dont la mesure extérieure-ϕ tend vers 0. En conséquence, si l'on retranche de P tous les ω joints à Γ_n, c'est-à-dire à l'un au moins des γ_i pour $i = 1, 2, \cdots, n$, la famille P_n des ω restants couvre indéfiniment, au sens de la métrique-ϕ, l'ensemble H_n. D'ailleurs, l'ensemble H'_n indéfiniment couvert par les P_n et qui est inclus dans H ne contient aucun point de Γ_n, puisque les ω de P_n sont tous disjoints de Γ_n. Donc H'_n est contenu dans $H - H \cdot \Gamma_n$. En conséquence H'_n contient H_n et est inclus dans $H - H \cdot \Gamma_n$; $H - H \cdot \Gamma_n$ excède H_n d'un ensemble inclus dans ou identique à Θ_n. Donc $H'_n = H_n + \Theta'_n$, Θ'_n étant une partie de Θ_n; Θ'_n peut exister, un même point pouvant être couvert indéfiniment (au sens

de la métrique-ϕ) à la fois par la famille des ω de P joints à Γ_n et par la famille des ω de P disjoints de Γ_n et qui forment P_n. En tous cas $\phi(\Theta'_n) = \phi(\Theta_n) = 0$.

Soit G_n l'ensemble des γ correspondant aux ω de P_n. Si γ est dans G_n, $\omega(\gamma)$ est disjoint de Γ_n. Comme γ est inclus dans $\omega(\gamma)$, *a fortiori* γ est disjoint de Γ_n, c'est-à-dire de tous les γ_i pour $i \leq n$.

Nous désignons par μ_{n+1} le maximum de $\phi(\gamma)$ pour les γ de G_n. Evidemment $\mu_{n+1} \leq \mu_n \leq \phi_e(D)$. Soit γ_{n+1} un des γ de G pour lequel $\phi(\gamma_{n+1}) > \mu_{n+1}/a$. Quel que soit γ dans G_n, $\phi(\gamma) < a\phi(\gamma_{n+1})$; γ_{n+1} étant dans G_n, est disjoint des γ_i antérieurs ($i \leq n$).

Ayant défini γ_{n+1}, nous sommes maintenant passés de Γ_n à Γ_{n+1}. Le mécanisme de la détermination progressive des γ_i est entièrement établi. Ces γ_i seront tous disjoints les uns des autres.

Les familles G_n décroissent, G_n contenant la totalité de G_{n+1} et en outre tout au moins l'ensemble γ_{n+1}.

13. Il reste à examiner le cas où la suite γ_i est illimitée, c'est-à-dire le cas où nous ne sommes pas arrêtés par une suite finie $\gamma_1, \gamma_2, \cdots, \gamma_n$ telle que $\phi(H - H \cdot \Gamma_n) = 0$.

Soit $\Gamma = \Sigma\gamma_i$. Nous voulons montrer l'égalité $\phi(H - H \cdot \Gamma) = 0$. Soit $\Theta = \Sigma\rho_i$; $\phi(\Theta) = 0$, puisque chacun des $\phi(\rho_i)$ est nul. Θ est dans H; $H - H \cdot \Gamma = R$ et $H - H \cdot (\Gamma + \Theta) = R'$ ont la même mesure extérieure. L'égalité $\phi(R) = 0$ équivaut à $\phi(R') = 0$. Prouvons celle-ci.

Observons que, les γ_i étant mesurables-ϕ et disjoints, Γ est mesurable-ϕ et $\phi(\Gamma) = \Sigma\phi(\gamma_i)$. L'inclusion $\Gamma \subset D$ résultant du fait que tout γ_i est dans D, entraîne $\phi(\Gamma) \leq \phi_e(D) < \infty$. Par suite, la série $\Sigma\phi(\gamma_i)$ converge et, en conséquence $\phi(\gamma_i)$ tend vers zéro, quand i croît.

Soit M un point de R'. Donnons-nous un nombre positif η quelconque et soit N un entier tel que $\sum_{n>N} \phi(\gamma_n) < \eta$.

M étant dans R' est inclus dans H et étranger à $\Sigma(\gamma_i + \rho_i)$. Donc M est étranger à $\gamma_i + \rho_i$ quel que soit i. Soit α (ou α'_N) le plus petit des nombres $\alpha_1, \cdots, \alpha_N$; M étant inclus dans H, il y a une infinité d'ensembles ω de mesure extérieure tendant vers 0 et contenant M. Pour les γ correspondants, $\phi(\gamma) < \phi_e(\omega)$ et $\phi(\gamma)$ tend aussi vers 0. Considérons un ω pour lequel $\gamma = \gamma(\omega)$ vérifie $\phi(\gamma) < \alpha$; ω est disjoint de $\gamma_1, \cdots, \gamma_N$; ω est donc dans P_N, et γ est dans G_N. Mais $\phi(\gamma_n)$ tendant vers 0 quand n croît, il y a un premier m tel que $a\phi(\gamma_{m+1}) \leq \phi(\gamma)$, tandis que $\phi(\gamma) < a\phi(\gamma_n)$ pour $n \leq m$; γ n'est donc pas dans G_m, et $m \geq N + 1$. Donce ω n'est pas dan P_m.

Les P_n, les G_n décroissent quand n croît. Il y a donc un i vérifiant $N+1 \leq i \leq m$, tel que γ est dans G_{i-1} (ensemble des γ tels que $\omega(\gamma)$ est disjoint de Γ_{i-1}) sans être dans G_i (ensemble des γ disjoints de Γ_i). Donc $\omega(\gamma)$ est joint à γ_i et, d'après $\phi(\gamma) < a\phi(\gamma_i)$; $(i \leq m)$, $\omega(\gamma)$ est dans l'ensemble $\Omega(\gamma_i)$, tout en contenant M. Donc M est dans $\Omega(\gamma_i)$. Soit $\Omega'_N = \sum_{n>N} \Omega(\gamma_n)$; M est dans Ω'_N ; M étant un point quelconque de R', il s'ensuit $R' \subset \Omega'_N$. Mais, en vertu de la deuxième hypothèse, $\phi_e[\Omega(\gamma_n)] < b\phi(\gamma_n)$. Donc

$$\phi_e[\Omega'_N] < b \sum_{n>N} \phi(\gamma_n) < b\eta.$$

Donc $\phi(R') < b\eta$ quel que soit η positif ; $\phi(R') = 0$ et $\phi(R) = 0$. La première conclusion du théorème est établie.

14. Passons à la seconde conclusion. Elle se décompose en deux parties. Montrons d'abord que H et Δ sont mesurables-ϕ.

Une conséquence évidente de la première conclusion du Théorème est que, Γ étant dans $D = \Sigma\gamma$, il en est de même de $H \cdot \Gamma$. Donc la partie de H étrangère à D est *a fortiori* étrangère à Γ. Elle est dans $R = H - H \cdot \Gamma$. C'est un ensemble de mesure-ϕ nulle.

η_n étant un nombre positif tendant vers 0 en décroissant quand n croît, soit D_n la réunion des γ vérifiant $\phi(\gamma) < \eta_n$. Si G^n est leur famille et si P^n est la famille des ω correspondant, P^n, qui comprend tous les ω pour lesquels $\phi_e(\omega) < \eta_n$, couvre indéfiniment, au sens de la métrique-ϕ, le même ensemble H que P elle-même. Les ensembles D_n, les familles G^n, P^n sont non croissants quand n croît.

D'après la première conclusion établie, on peut couvrir H, exception faite d'un ensemble R_n de mesure-ϕ nulle, avec une suite d'ensembles $\gamma_1^n, \gamma_2^n, \cdots, \gamma_i^n, \cdots$ disjoints, appartenant à la famille G^n, donc situés dans D_n. Si $\Gamma_n = \sum_i \gamma_i^n$, l'ensemble $R_n = H - H \cdot \Gamma_n$ vérifie $\phi(R_n) = 0$.

Soit Δ l'ensemble ΠD_n, commun à tous les D_n ; Δ est l'ensemble indéfiniment couvert par la famille G, au sens de la métrique-ϕ. Car tout point indéfiniment couvert par la famille G au sens de la métrique-ϕ appartient à une suite de γ tels que $\phi(\gamma)$ tende vers 0. Ce point appartient donc à tous les D_n. Il est donc dans Δ. Réciproquement, tout point de Δ, étant commun à tous les D_n, appartient à un γ pour lequel $\phi(\gamma) < \eta_n$, quel que soit n. Ce point est couvert indéfiniment par la famille G, au sens de la métrique-ϕ.

Evidemment $\Delta \subset H$, puisque un point couvert indéfiniment (au sens de la métrique-ϕ) par les γ l'est *a fortiori* par les $\omega(\gamma)$ correspondants. En effet, d'une part $\omega(\gamma)$ contient γ, d'autre part d'après $\phi_e[\omega(\gamma)] < b\phi(\gamma)$ (consé-

quence de la seconde hypothèse), $\phi_e[\omega(\gamma)]$ tend vers 0 quand il en est ainsi de $\phi(\gamma)$.

Soit $\Delta' = \Pi\Gamma_n$. Tous les Γ_n sont mesurables-ϕ. Donc Δ' est également mesurable-ϕ; Δ' est inclus dans Δ. Car, Γ_n étant dans D_n, $\Pi\Gamma_n = \Delta'$ est dans $\Pi D_n = \Delta$. En conséquence

$$\Delta' \subset \Delta \subset H.$$

D'ailleurs $H - \Delta' \subset \Sigma R_n$. Car si un point de H n'est pas dans Δ', il n'est pas commun à tous les Γ_m. It est étranger à au moins un Γ_m. Etant dans H, il est dans R_m. Donc l'ensemble $S' = H - \Delta'$ est compris dans ΣR_n. Or $\phi(\Sigma R_n) = 0$. Donc $H = \Delta' + S'$ avec $\phi(S') = 0$.

Δ' étant mesurable-ϕ, il en est de même de H et les deux ensembles H et Δ' ont la même mesure. En vertu des inclusions $\Delta' \subset \Delta \subset H$, Δ est lui aussi mesurable-ϕ et a même mesure que H et que Δ'. Si $H = \Delta + S$, $S \subset S'$ et $\phi(S) = 0$.

15. La première partie de la seconde conclusion du Théorème est établie. Passons à la seconde partie, à savoir la possibilité de choisir les γ_i de façon que $\phi(\Delta) \leq \phi(\Gamma) < \phi(\Delta) + \epsilon$, le nombre positif ϵ étant préalablement donné.

D'abord la condition $\phi(\Delta) \leq \phi(\Gamma)$ est vérifiée par toute collection $\Gamma = \Sigma\gamma_i$ telle que $\phi(H - H \cdot \Gamma) = 0$. Car, d'après $H = \Delta + S$ et $\phi(S) = 0$, $\Gamma \cdot H = \Gamma \cdot \Delta + \Gamma \cdot S$, avec $\phi(\Gamma \cdot S) = 0$. Donc $\phi(\Gamma \cdot \Delta) = \phi(\Gamma \cdot H)$. Or $\phi(\Gamma \cdot H) = \phi(H)$, $\phi(H) = \phi(\Delta)$, et évidemment $\phi(\Gamma) \geq \phi(\Gamma \cdot \Delta)$. Donc $\phi(\Gamma) \geq \phi(\Delta)$.

Reste à prouver la possibilité que $\phi(\Gamma) < \phi(\Delta) + \epsilon$ pour un choix approprié des γ_i dont la collection est Γ. Evidemment, il suffit pour cela d'établir l'égalité $\lim \phi(\Gamma_n) = \phi(\Delta)$, la collection Γ_n étant celle du paragraphe précédent, composée d'ensembles γ_i^n de G^n, famille composée des γ vérifiant $\phi(\gamma) < \eta_n$.

D_n est la réunion de ces mêmes γ composant G_n, et Γ_n est donc inclus dans D_n; D_n décroît (plus précisément: ne croît pas) quand n croît, et l'ensemble-limite abstraite de D_n (ensemble commun à tous les D_n, parce que ces ensembles décroissent) est Δ.

Si D_n est mesurable-ϕ, tout au moins à partir d'une certaine valeur de n, $\phi(D_n)$ tend vers $\phi(\Delta)$. D'après $\phi(\Delta) \leq \phi(\Gamma_n) \leq \phi(D_n)$, il est évident que $\phi(\Gamma_n)$ tend vers $\phi(\Delta)$.

Le point à établir ne présente de difficulté apparente que si l'on ne suppose pas D_n mesurable-ϕ, ce que *a priori* ne peut être écarté.

Le fait que $\Delta = \lim D_n$ soit mesurable-ϕ, les D_n étant des ensembles décroissants, n'implique nullement ni que D_n soit mesurable-ϕ, ni que $\phi_e(D_n)$ tende vers $\phi(\Delta)$. Il est aisé de former un exemple de cette dernière circonstance, et avec la mesure euclidienne. On sait que l'on peut décomposer l'intervalle $\sigma(0,1)$ des nombres réels en une infinité dénombrable d'ensembles E_n disjoints et dont les longueurs extérieures $l_e(E_n)$ sont toutes égales à 1, leurs longueurs intérieures étant dès lors toutes nulles.[5] Posons $\sigma_n = \sum_{m \geq n} E_m$, en sorte que $\sigma = \sigma_1$. Dès $n \geq 2$, σ_n, ayant comme chacun des E_m, pour $m \geq n$, la mesure extérieure 1, tandis que son complémentaire $\sigma - \sigma_n$, a lui aussi, comme chacun des E_i pour $i < n$ la même mesure extérieure 1, σ_n est non mesurable. Quand n croît, σ_n décroît, et tout point de σ cesse à partir d'une certaine valeur de n de faire partie de σ_n, puisque ce point appartient à un des E_i et à un seul. Donc l'ensemble commun aux σ_n, soit $\lim \sigma_n$, est vide. Or $l_e(\sigma_n) = 1$ indépendamment de n; $l_e(\sigma_n)$ ne tend pas vers $l(\lim \sigma_n) = 0$.

---

[5] Voir par exemple Denjoy, *Leçons sur le calcul des coefficients des séries trigonométriques*, 2e partie, pp. 146-147.

Le passage cité appelle toutefois un développement complémentaire pour permettre d'affirmer que les E_n, tous situés sur σ, ont tous pour mesure extérieure l'unité. Rappelons les considérations initiales.

On énumère en une suite unique $\theta_1, \theta_2, \cdots, \theta_n, \cdots$ les nombres rationnels de l'intervalle $\sigma(0,1)$. Soit η leur ensemble. x étant un point quelconque de σ, désignons par $e(x)$ l'ensemble formé de x et des points x' de σ tels que $x' - x$ $(x < x')$ ou $x' - x + 1$ $(x' < x)$ soit dans η; $e(x + \theta_n)$ si $x + \theta_n < 1$, $e(x + \theta_n - 1)$ si $x + \theta_n > 1$ sont identiques à $e(x)$. Soit E_0 un ensemble contenant un point et un seul de chacun des ensembles $e(x)$, et E_n l'ensemble des $x + \theta_n$ $(x + \theta_n < 1)$ ou $x + \theta_n - 1$ $(x + \theta_n > 1)$, x décrivant E_0. Les E_n sont disjoints. Ils ont tous la mesure intérieure zéro, la même mesure extérieure β positive. Seulement *a priori*, β n'est pas nécessairement égal · 1 (β peut être supposé aussi petit que l'on veut, E_0 pouvant être placé sur l'intervalle $0, \beta$).

Il existe sur σ un ensemble G_δ de mesure β contenant E_0 et, si $0 < \lambda < \beta$, il y a dans G_δ un ensemble parfait P, épais en lui-même, de mesure λ. L'ensemble $E_0 \cdot P = I_0$ a pour mesure extérieure λ (sans quoi E_0 n'aurait pas pour mesure extérieure β). Soit $u_1, u_2, \cdots, u_k, \cdots$ la suite des intervalles contigus à P, rangés par longueurs non croissantes. Le nombre $s_k = k u_k + \sum_{p \geq k+1} u_p$ tend vers 0 si k augmente. Imprimons à P une translation rationnelle θ_n inférieure à u_k. L'ensemble parfait obtenu P_k est tel que $P \cdot P_k$ a une longueur supérieure à $\lambda - s_k$. Soit $k_1, k_2, \cdots, k_m, \cdots$ une suite d'entiers croissants tels que la série $\sum s_{k_m}$ converge et ait une somme inférieure à λ. Si P_m et I_m sont respectivement déduits de P et de I_0 par une translation θ_{n_m} inférieure à u_{k_m}, d'une part $I_m = E_{n_m} \cdot P_m$, d'autre part λ est à la fois la longueur de P_m et la longueur extérieure de I_m. Si Q est l'ensemble majeur épais en lui-même commun aux P_m, la longueur μ de Q est positive et Q est parfait. Les ensembles $J_m = I_m \cdot Q = E_{n_m} \cdot Q$ sont disjoints et ont la mesure extérieure μ. Si a est l'origine de Q et $m(x)$ la longueur de Q entre a et x $(> a)$, la transformation $y = m(x)/\mu$ change Q en le segment $(0,1)$ et les ensembles J_m en des ensembles H_m situés sur σ, disjoints et tous de mesure intérieure zéro, de mesure extérieure 1.

16. Supposons que D_n soit non mesurable-ϕ; $D_n - \Delta$ est non mesurable-ϕ. Cet ensemble peut se décomposer en deux ensembles disjoints (le premier éventuellement vide) X_n et Y_n tels que: 1° X_n est mesurable-ϕ et $\phi(X_n)$ a la plus grande valeur possible, en sorte que $\phi(X_n) = \phi_i(D_n - \Delta)$; 2° Y_n est non mesurable-ϕ, sa mesure intérieure $\phi_i(Y_n)$ étant nulle d'après la valeur de $\phi(X_n)$; D_n est décomposé en trois ensembles disjoints $D_n = \Delta + X_n + Y_n$ et $\phi_e(D_n) = \phi(\Delta) + \phi(X_n) + \phi_e(Y_n)$, $\phi_i(D_n) = \phi(\Delta) + \phi(X_n)$.

On peut supposer que X_n décroît (ne croît pas) quand n croît. Car X_{n+1} inclus dans D_{n+1} l'est *a fortiori* dans D_n. Les points de X_{n+1} étrangers à X_n forment un ensemble mesurable-ϕ dont la mesure-ϕ est nulle. Sans quoi, en ajoutant ces points à X_n, on augmenterait $\phi(X_n)$, sans que X_n cessât d'être inclus dans D_n, ce qui est impossible, puisque $\phi(X_n) = \phi_i(D_n - \Delta)$ est par hypothèse le maximum des mesures-ϕ des ensembles mesurables-ϕ inclus dans $D_n - \Delta$. En retranchant de X_{n+1} les points étrangers à X_n pour les ajouter à Y_{n+1}, on n'altere pas la condition $\phi(X_{n+1}) = \phi_i(D_{n+1} - \Delta)$, et l'on réalise l'inclusion $X_{n+1} \subseteq X_n$.

X_n décroît. Comme $X_n + Y_n$ égal à $D_n - \Delta$ décroissant (non croissant) a pour limite abstraite (ensemble commun) l'ensemble vide, pareillement $\lim X_n$ est vide, et $\phi(X_n)$ tend vers 0 quand n croît.

Or, Γ_n étant dans D_n, $\Gamma_n = \Gamma_n \cdot D_n = \Gamma_n \cdot \Delta + \Gamma_n \cdot X_n + \Gamma_n \cdot Y_n$; Γ_n est mesurable-ϕ, comme somme des γ_i eux-mêmes mesurables-ϕ par hypothèse. De même, $\Gamma_n \cdot \Delta$ et $\Gamma_n \cdot X_n$ sont mesurables-ϕ puisque Δ et X_n le sont. Donc $\Gamma_n \cdot Y_n$ est mesurable-ϕ, et comme cet ensemble est inclus dans Y_n dont la mesure intérieure-ϕ est nulle, $\phi(\Gamma_n \cdot Y_n) = 0$. En conséquence $\phi(\Gamma_n) = \phi(\Gamma_n \cdot \Delta) + \phi(\Gamma_n \cdot X_n)$.

$\phi(\Gamma_n \cdot X_n)$ tend vers 0 puisqu'il en est ainsi de $\phi(X_n)$; $\phi(\Gamma_n \cdot \Delta) = \phi(\Delta)$, indépendamment de n. Donc $\lim \phi(\Gamma_n) = \phi(\Delta)$.

ϵ étant un nombre positif donné, il est donc possible de choisir η_n assez petit pour que la collection Γ_n ou Γ composée avec des ensembles disjoints γ_i de la famille définie par $\phi(\gamma) < \eta_n$ vérifie $\phi(\Gamma) < \phi(\Delta) + \epsilon$.

La seconde conclusion du Théorème Général est entièrement démontrée. La proposition complète est donc établie.

17. Sur les trois hypothèses conditionnant le théorème général, on peut se demander si la troisième $[\phi_e(\Sigma\gamma) < \infty]$ est indispensable. Un exemple réalisé dans le plan U_2, avec la métrique de l'aire euclidienne des ensembles, va nous montrer un couple de familles P, G d'ensembles ω et γ respectivement, vérifiant les deux premières hypothèses du théorème, l'ensemble H indéfiniment recouvert, au sens des aires, par la famille G des ensembles ω ayant une

aire déterminée et finie, tandis que la conclusion du théorème général est mise en défaut. Non seulement H ne peut pas être couvert par une collection dénombrable d'ensembles γ, mais les γ sont tous disjoints de H.

Regardons U_2 comme le plan complexe de la variable $z = x + iy = re^{i\theta}$. Soit C le cercle ouvert $r < 1$; $\epsilon_1, \epsilon_2, \cdots, \epsilon_n, \cdots$ étant une suite de nombres positifs décroissants tendant vers 0, avec $\epsilon_1 < 1/4$, nous formons une série de familles associées, P_n d'ensembles ω et G_n d'ensembles γ, de la façon suivante. $z = re^{i\theta}$ étant un point quelconque de C, nous constituons un ensemble ω de P_n par les deux cercles de même rayon ϵ_n et de centres respectifs z et $z + 2ne^{i\theta} = (r + 2n)e^{i\theta}$. L'ensemble $\gamma = \gamma(\omega)$ correspondant dans G_n sera le second de ces cercles. Les familles P et G sont respectivement ΣP_n et ΣG_n.

Les cercles γ de G_n forment par leur réunion la couronne circulaire I_n $(2n - \epsilon_n < r < 2n + 1 + \epsilon_n)$. La distance de deux de ces couronnes consécutives et celle de I_1 à I_0 $(0 < r < 1 + \epsilon_1)$ surpassent 1/2.

L'ensemble H indéfiniment couvert par la famille G, au sens des aires (et de la dimension, et au sens de toute métrique donnant la mesure zéro à tout point et une mesure positive à tout ensemble ouvert) est le cercle fermé $\bar{C}$ $(r \leq 1)$. L'ensemble $D = \Sigma\gamma$ est $\sum_{n \geq 1} I_n$; D est distant de H d'au moins 3/4. Telle est la distance minimum de tout ensemble γ à H.

Or les deux premières hypothèses du théorème général sont vérifiées par les familles P, G. En effet, si un ensemble $\omega' = \omega(\gamma')$ est joint à un γ et si celui-ci est dans G_n, ω' est joint à γ par sa partie $\gamma(\omega')$, uniquement, puisque $\omega' - \gamma(\omega')$ est dans I_0. Donc $\gamma(\omega')$ est dans I_n, comme γ; ω' est dans P_n comme $\omega(\gamma)$. Il suit de là que, les ω' joints à γ, ayant leur aire bornée inférieurement par $2\pi\epsilon_n^2$, ne recouvrent indéfiniment aucun point, au sens des aires (ni d'ailleurs au sens d'aucune métrique donnant une mesure positive à tout ensemble ouvert). Donc $\rho(\gamma)$ est vide quel soit γ. La première hypothèse $\phi[\rho(\gamma)] = 0$ est vérifiée.

[Pour associer aux conclusions précédentes une métrique borélienne $\phi(E)$ plus générale que celle des aires euclidiennes, il conviendrait d'exiger plus précisément que, si c est un cercle de rayon r, $\phi(c)$ vérifie une double inégalité $0 < h(r) < \phi(c) < k(r)$, h et k étant indépendants de c et $k(r)$ tendant vers 0 avec r].

Passons à la seconde. Quel que soit $a > 1$, les ω' vérifiant $\phi(\gamma') < a\phi(\gamma)$, si $\gamma' = \gamma(\omega')$, et joints à γ sont, en raison de cette dernière condition, dans P_n. Ils s'identifient à tous les ω' joints à γ. Pour eux $\phi(\gamma') = \phi(\gamma)$, avec la métrique-$\phi$ des aires. Chacun de ces ω' se compose de deux cercles de même rayon ϵ_n que γ. Leur réunion $\Omega(\gamma)$ est contenue dans un couple de

cercles ayant le même rayon $3\epsilon_n$ et respectivement concentriques aux deux cercles composant $\omega(\gamma)$. Donc $\Omega(\gamma)$ a son aire inférieure à 18 fois l'aire de γ. La seconde hypothèse est vérifiée.

Mais la troisième: aire $D < \infty$ ne l'est pas puisque $D = \Sigma I_n$, l'aire de I_n surpassant $4\pi n$. Cette troisième hypothèse joue donc un rôle indispensable dans le théorème général.

18. Un cas particulier important est celui où les ensembles ω se confondent avec les ensembles γ. C'est par l'examen de cette hypothèse que nous terminerons cette étude.

Nous dirons que, dans l'espace U où a été définie une métrique $\phi(E)$ des ensembles, borélienne, non négative, une *famille* G d'ensembles γ est *régulière au sens de la métrique-ϕ* si:

1° *Tout ensemble γ a une mesure-ϕ déterminée positive et finie;*

2°. *Quel que soit l'ensemble γ de G, les points étrangers à γ et couverts indéfiniment, au sens de la métrique-ϕ, par les γ' joints à γ forment un ensemble $\rho(\gamma)$ de mesure-ϕ nulle;*

3°. *Il existe deux nombres a et b $(1 < a < b)$ indépendants de γ, tels que la réunion $\Omega(\gamma)$ des γ' joints à γ et vérifiant $\phi(\gamma') < a\phi(\gamma)$ a une mesure extérieure-ϕ inférieure à $b\phi(\gamma)$: $\phi_e[\Omega(\gamma)] < b\phi(\gamma)$;*

4°. *La réunion des γ est un ensemble D de mesure extérieure finie: $\phi_e(D) < \infty$.*

Nous désignons par $\Delta(G)$ l'*ensemble indéfiniment couvert, au sens de la métrique-ϕ, par la famille G des ensembles γ.*

Le théorème général (10), où l'on identifie les ω avec les γ, donne l'énoncé suivant:

THÉORÈME. *L'ensemble $\Delta(G)$ relatif à la famille G d'ensembles γ, régulière au sens de la métrique-ϕ, peut être couvert, à un ensemble de mesure-ϕ nulle près, par une collection dénombrable $\Gamma = \Sigma\gamma_i$ d'ensembles $\gamma_1, \gamma_2, \cdots$, disjoints, appartenant à G, et $\Delta(G)$ est mesurable-ϕ.* En outre, quelque soit le nombre *positif* ϵ préalablement donné, *on peut toujours réaliser l'égalité* $\phi(\Gamma) < \phi(\Delta) + \epsilon$.

Il est à peine besoin de noter que, dans l'espace cartésien U_r à r dimensions et si la métrique $\phi(E)$ est la mesure euclidienne borélienne d'ordre r, on a une famille régulière dans tout groupement, borné dans U_r, d'ensembles γ ayant une mesure déterminée d'ordre r, si cette dernière surpasse une

fraction fixe (indépendante de γ) de d^r, d étant le diamètre de γ. Par exemple, les γ peuvent être des sphères, ou des quasi-parallèlipipèdes rectangles d'ordre r, dont les côtés ont des rapports mutuels compris entre deux nombres fixes k et $1/k$ $(0 < k < 1)$. C'est à cette dernière sorte d'ensembles que se rapporte le théorème classique de Vitali. Mais notre définition des familles régulières englobe évidemment des espèces d'ensembles bien plus générales, où la régularité de l'assemblage doit être assurée dans chaque cas particulier par la consistance de la famille considérée.

19. Pour le dernier théorème (18) on peut se poser la même question que pour le Théorème général: Si les ensembles γ de la famille G vérifient les deux premières hypothèses, et si l'on suppose en outre que l'ensemble $\Delta(G)$ indéfiniment couvert par les γ a une mesure extérieure finie, la troisième hypothèse est-elle ou non superflue?

Dans le plan U_2 et avec la mesure euclidienne des aires, soit C_0 le carré ouvert $(0 < x < 1, 0 < y < 1)$. Divisons-le en quatre carrés égaux, et chacun de ceux-ci en neuf autres dont les carrés centraux fermés seront notés C_1^1. Nous avons partagé C_0 en 36 carrés égaux dont 4 ont été retenus. Il en reste 32, se partageant en quatre groupes de huit entourant un même carré C_1^1. Dans chacun de ces 32 carrés complémentaires opérons comme nous l'avons fait sur C_0: division en 36 carrés égaux sur lesquels 4 sont retenus et notés C_1^2. Nous répétons indéfiniment l'opération. Nous avons des collections de 4 carrés C_1^1, 4. 32 carrés C_1^2, 4. 32^2 carrés $C_1^3, \cdots$, 4. 32^{n-1} carrés C_1^n, etc. Les carrés C_1^n sont disjoints et leur réunion $\Sigma C_1^n = \Lambda^1$ couvre C_0 moins un ensemble d'aire nulle. Car l'aire de C_0 diminué des carrés C_1^1 vaut 8/9, l'aire de C_0 moins les C_1^1 et les C_1^2 vaut $(8/9)^2, \cdots$. L'aire de $C_0 - \Sigma C_1^1 - \Sigma C_1^2 \cdots - \Sigma C_1^n$ vaut $(8/9)^n$ et tend vers 0 quand n croît.

Maintenant, sur chacun des C_1^n $(n = 1, 2, \cdots)$ procédons comme sur C_0. Nous obtenons une famille Λ_2 de carrés fermés disjoints C_2^p qui couvrent chacun des C_1^n à un ensemble d'aire nulle près et C_0 de la même manière. Et nous recommençons avec les C_2^p pour obtenir une famille Λ_3 de carrés C_3^q disjoints, couvrant C_0 à un ensemble d'aire nulle près, etc. La réunion des familles Λ_m est une famille Λ d'ensembles fermés C_m^k couvrant indéfiniment un ensemble Υ qui est C_0 diminué d'un ensemble d'aire nulle (formé des points se projetant sur l'un ou l'autre de deux ensembles égaux, sommes d'ensembles parfaits totalement discontinus privés de leurs points de première espèce et situés respectivement sur Ox et sur Oy).

La question qui se pose est celle-ci. Peut-on adjoindre à chacun des C_m^k

un ensemble g_m^k extérieur à C_0, et par exemple fermé, la réunion $C_m^k + g_m^k$ formant un ensemble γ (ou $\gamma_{m,k}$), de façon que, ces trois conditions étant vérifiées :

1°. l'ensemble Δ indéfiniment couvert, au sens des aires, par la famille G des γ est tout entier dans C_0, donc identique à Υ (l'ensemble indéfiniment couvert par la famille des g_m^k est vide) ;

2° et 3°. les γ vérifient les deux premières hypothèses,

néanmoins il soit impossible de couvrir $\Delta \equiv \Upsilon$ (moins un ensemble d'aire nulle) avec une infinité dénombrable de γ_i disjoints, parce que, pour aucun choix d'une collection de C_m^k disjoints et couvrant C_0 sauf un ensemble d'aire nulle, les g_m^k correspondants ne sont tous disjoints ?

Bien entendu cette impossibilité exige que l'aire de $D = \Sigma\gamma$ (donc celle de Σg_m^k) soit infinie.

Je laisse au lecteur le soin d'élucider ce point.

20. Il est évident que *toute famille partielle d'une famille régulière G est elle-même régulière*, toujours au sens de la métrique-ϕ.

On peut se poser la question suivante. Si $\Delta(G)$ est décomposé en deux ensembles mesurables-ϕ disjoints, H' et H'', est-il toujours possible de trouver dans la famille G deux familles disjointes G' et G'' d'ensembles γ de façon que $\Delta(G') = H'$, à un ensemble de mesure-ϕ nulle près, et sous la même réserve $\Delta(G'') = H''$?

Il n'en est rien, en voici un exemple. Dans le plan U_2 où la métrique est celle des aires boréliennes, soient C_1 et C_2 deux couronnes circulaires ouvertes, homothétiques par rapport à leur centre commun 0 et disjointes, par exemple C_1 de rayon 1 et 2, C_2 de rayons 3 et 6. Prenons pour ensembles γ tous les couples de cercles c_1, c_2, homothétiques entre eux avec le rapport de 1 à 3, par rapport à 0, et respectivement intérieurs à C_1 et à C_2. On voit immédiatement que la famille G des γ vérifie les quatre conditions de régularité. Si les deux ensembles $\gamma = c_1 + c_2$ et $\gamma' = c'_1 + c'_2$ de G sont joints, simultanément c_1 et c'_1 d'une part, c_2 et c'_2 d'autre part sont joints. D'où résultent immédiatement les second et troisième caractères de régularité de la famille G des γ. Le premier et le quatrième caractères sont évidemment réalisés aussi. $\Delta(G)$ est identique à $C_1 + C_2$.

Toute famille partielle de G couvre deux ensembles distincts, l'un situé dans C_1, l'autre dans C_2, le second homothétique du premier par rapport à 0, avec le rapport 3. Donc, quel que soit l'ensemble H_1 dans C_1, si son homothétique par rapport à 0 avec le rapport 3 est l'ensemble H_2 situé

dans C_2, toute famille partielle de G couvrant indéfiniment H_1 couvre en même temps H_2. Et réciproquement. En particulier, si $\Delta(G) = C_1 + C_2$ est décomposé en $H' = C_1$ et $H'' = C_2$, toute famille G' partielle de G couvrant indéfiniment H' et toute famille G'' partielle de G couvrant indéfiniment H'', toujours au sens des aires, et à un ensemble d'aire nulle près, donnent lieu aux identités: $\Delta(G') = \Delta(G'') = \Delta(G) = C_1 + C_2$, à un ensemble d'aire nulle près.

Cet exemple, qui pourrait être transformé dans le détail (les γ de G composés de couples de cercles c_1, c_2 homothétiques avec le rapport 3 ne couvrant que deux régions homothétiques de C_1 et de C_2) fait saisir pourquoi, contrairement à ce que l'on pourrait supposer, les conclusions des théorèmes qui suivent n'admettent pas en général la substitution aux ensembles $\Delta(G)$ de n'importe laquelle de leurs parties mesurables.

21. Ce sont les familles régulières G telles que les précédentes qui donnent de l'intérêt à la proposition suivante:

THÉORÈME. *Au sens de la métrique-ϕ, si l'ensemble H de mesure extérieure-ϕ positive est inclus dans l'ensemble $\Delta = \Delta(G)$ indéfiniment couvert par la famille régulière G d'ensembles γ, et si H vérifie, pour tout ensemble γ de G l'inégalité $\phi_e(H \cdot \gamma) \prec \theta\phi(\gamma)$, θ étant inférieur à 1 et indépendant de γ, l'inégalité $\phi(\Delta) < \theta^{-1}\phi_e(H)$ est impossible.*

Soit en effet, $\Gamma = \Sigma\gamma_i$ l'ensemble-somme d'une collection dénombrable d'ensembles $\gamma_1, \gamma_2, \cdots$ de G, disjoints et couvrant Δ à un ensemble de mesure-ϕ nulle près. En outre, ϵ désignant (et nous ne le répèterons plus dans les pages suivantes) un nombre positif indépendant, on peut réaliser la condition $\phi(\Gamma) < \phi(\Delta) + \epsilon$ (15-16).

De l'égalité $\Delta = \Delta \cdot \Gamma + \Theta$ avec $\phi(\Theta) = 0$, et de la relation d'inclusion $H \subset \Delta$, il résulte: $H = H \cdot \Delta = H \cdot \Gamma + H \cdot \Theta$ et $\phi(H \cdot \Theta) = 0$. Donc $\phi_e(H) = \phi_e(H \cdot \Gamma)$. Les ensembles γ_i étant disjoints (6), $\phi_e(H \cdot \Gamma) = \Sigma\phi_e(H \cdot \gamma_i)$. Donc $\phi_e(H) < \theta\Sigma\phi(\gamma_i) = \theta\phi(\Gamma) < \theta[\phi(\Delta) + \epsilon]$; ϵ étant positif quelconque, $\phi_e(H) \leq \theta\phi(\Delta)$. L'inégalité $\phi(\Delta) < \theta^{-1}\phi_e(H)$ est impossible.

22. Les espaces généraux U, avec les métriques quelconques, donnent lieu à un théorème coincidant pour les espaces cartésiens et la métrique euclidienne borélienne, les familles G étant composées de sphères, avec le théorème sur l'*épaisseur* des ensembles mesurables.[6]

---

[6] J'emploie le terme d'*épaisseur* où trop d'analystes usent de celui de *densité*. Je

Dans U soit G une famille d'ensembles γ. Au sens de la métrique-ϕ, supposons G régulière et couvrant indéfiniment l'ensemble $\Delta = \Delta(G)$. Soit E un ensemble mesurable-ϕ inclus dans Δ, M un point de Δ et $\eta_i(E, M)$, $\eta_s(E, M)$ respectivement *la plus petite* et *la plus grande limites du rapport* $\phi(E \cdot \gamma)/\phi(\gamma)$ *quand l'ensemble* γ *de* G *contient* M et *que sa mesure* $\phi(\gamma)$ *tend vers* 0; $\eta_i(E, M)$, $\eta_s(E, M)$ seront appelées respectivement *l'épaisseur inférieure, l'épaisseur supérieure de l'ensemble* E *au point* M, *relativement à la famille* G. Si ces deux nombres sont égaux, leur valeur commune sera appelée *l'épaisseur de* E *au point* M *relativement à la famille* G.

THÉORÈME. *Au sens de la métrique-*ϕ, *si la famille* G *d'ensembles* γ *est régulière et si* $\Delta = \Delta(G)$ *est l'ensemble indéfiniment couvert par elle, l'épaisseur de* Δ *relativement à la famille* G *est* 1 *en tout point de* Δ, *sauf éventuellement sur un ensemble de mesure-*ϕ *nulle.*

L'hypothèse contraire est que l'ensemble des points M vérifiant $\eta_i(\Delta, M) < 1$ a une mesure extérieure-ϕ positive. Considérons l'ensemble des M vérifiant $\eta_i(\Delta, M) < 1 - 1/n$. Si, quel que fût n, cet ensemble avait la mesure-ϕ nulle, l'ensemble $\eta_i(\Delta, M) < 1$ aurait aussi la mesure 0. Il y a donc un nombre $\theta < 1$ tel que l'ensemble H des points M vérifiant $\eta_i(\Delta, M) < \theta$ a sa mesure extérieure $\phi_e(H)$ positive.

Soit G' la famille des ensembles γ appartenant à G et vérifiant $\phi(\Delta \cdot \gamma) < \theta\phi(\gamma)$. L'ensemble $\Delta' = \Delta(G')$, indéfiniment couvert par la famille G' contient H. On peut d'ailleurs observer qu'en tout point de $\Delta' - H$, on trouve $\eta_i(\Delta, M) = \theta$; Δ' est mesurable-ϕ (14). Donc $\phi(\Delta') \geq \phi_e(H) > 0$. Couvrons Δ', à un ensemble de mesure-ϕ nulle près, avec une collection dénombrable d'ensembles $\gamma'_1, \gamma'_2, \cdots$ de G', disjoints, et soit $\Gamma' = \Sigma\gamma'_i$. On peut supposer $\phi(\Gamma') < \phi(\Delta') + \epsilon$. (15-16). D'après $\Delta' = \Delta' \cdot \Gamma' + \Theta'$, avec $\phi(\Theta') = 0$, $\phi(\Delta') = \phi(\Delta' \cdot \Gamma')$. L'inclusion $\Delta' \subset \Delta$ donne $\Delta' \cdot \Gamma' \subset \Delta \cdot \Gamma'$ et $\phi(\Delta') \leq \phi(\Delta \cdot \Gamma')$. Mais, les γ'_i étant disjoints

$$\phi(\Delta \cdot \Gamma') = \Sigma\phi(\Delta \cdot \gamma'_i) < \theta\Sigma\phi(\gamma'_i) = \theta\phi(\Gamma') < \theta[\phi(\Delta') + \epsilon].$$

En conséquence $\phi(\Delta') < \theta[\phi(\Delta') + \epsilon]$, quel que soit ϵ positif. Donc

$$\phi(\Delta') \leq \theta\phi(\Delta'),$$

---

réserve celui-ci à la notion topologique, fondamentale dans la théorie des fonctions de variables réelles.

Voir mes *Leçons sur le Calcul des coefficients des séries trigonométriques* (note liminaire p. XI, pp. 105-106 en note; p. 223 en note). Au sujet de la densité, voir également: p. 136 en note, 331 en note. Et consulter la Critique du Langage, pp. 689-690.

ce qui est absurde si, avec $\theta < 1$, en même temps $\phi(\Delta') > 0$. Or il en est ainsi au cas où $\phi_e(H) > 0$.

23. APPLICATIONS AUX FONCTIONS D'ENSEMBLES. Dans l'espace U pourvu de la métrique-ϕ, soit $\psi(E)$ une fonction d'ensemble complètement additive, et même au sens précis suivant: Si les ensembles $E_1, E_2, \cdots$ sont disjoints, si tous les $\psi(E_n)$ sont déterminés et si la série $\Sigma\,|\,\psi(E_n)\,|$ converge, $\psi(E)$ existe et vaut $\Sigma\psi(E_n)$. En outre, on suppose la soustractivité de $\psi(E)$: Si, avec $E \supset E'$, $\psi(E)$ et $\psi(E')$ sont définis, $\psi(E - E')$ existe et dès lors vaut $\psi(E) - \psi(E')$.

Enfin (6), dès que $\psi(E)$ et $\psi(E')$ sont définis, $\psi(E \cdot E')$ l'est aussi.

Dans U soit G une famille d'ensembles γ. Au sens de la métrique-ϕ, nous supposons G régulière et couvrant indéfiniment l'ensemble $\Delta = \Delta(G)$; Δ est mesurable-ϕ (14). Nous dirons que la fonction $\psi(E)$ est *définie dans le champ* $\Delta(G)$ si $\psi(E)$ est déterminé et fini pour tout ensemble H mesurable-ϕ inclus dans Δ (y compris Δ). Nous ne rappellerons plus désormais les hypothèses générales ni le sens des notations adoptées.

Soit M un point quelconque de Δ. Nous appelerons *nombres dérivés supérieur, inférieur, moyen, relatifs à la famille G et propres à la fonction d'ensemble* ψ *au point M*, respectivement la plus grande, la plus petite, une quelconque des limites du rapport $\psi(\Delta \cdot \gamma)/\phi(\gamma)$ si l'ensemble γ de G contient M et si $\phi(\gamma)$ tend vers 0.

Nous noterons ces nombres $\Lambda_s(\psi, M, G)$, $\Lambda_i(\psi, M, G)$, $\Lambda_m(\psi, M, G)$.

Nous dirons que $\psi(E)$ possède *au point M une dérivée relative à la famille régulière G* si $\Lambda_s(\psi, M, G) = \Lambda_i(\psi, M, G)$. Cette dérivée sera notée $\Lambda(\psi, M, G) = \Lambda_i = \Lambda_s$.

Je dis que *les nombres dérivée* Λ_i *et* Λ_s *sont mesurables-*ϕ. Une fonction de point $f(M)$ est dite *mesurable-*ϕ si, H désignant son champ de définition dans U, l'ensemble des M où $A < f(M) < B$ (M inclus dans H) est mesurable-ϕ, quels que soient A et B ($A < B$).

Soit A un nombre réel quelconque et I_A l'ensemble où $\Lambda_i(\psi, M, G) < A$. Si nous formons la famille $G_i(A)$ des ensembles γ de G pour lesquels $\psi(\Delta \cdot \gamma) < A\phi(\gamma)$, l'ensemble $\Delta[G_i(A)]$ indéfiniment couvert par cette famille renferme I_A et les points de $\Delta[G_i(A)]$ étrangers à A vérifient $\Lambda_i(\psi, M, G) = A$.

Or cet ensemble $\Delta[G_i(A)]$ est mesurable-ϕ, puisque la famille $G_i(A)$, partie de la famille régulière G, est elle-même régulière au sens de la métrique-ϕ. Donc, quel que soit A, il y a un ensemble I'_A, savoir $\Delta[G_i(A)]$, mesurable-ϕ et qui d'une part contient l'ensemble $\Lambda_i(\psi, M, G) < A$, d'autre

part est compris dans l'ensemble $\Lambda_i(\psi, M, G) \leq A$. On en conclut, parle raisonnement de Lebesgue relatif aux fonctions mesurables d'une variable réelle, que la fonction $\Lambda_i(\psi, M, G)$ est mesurable-ϕ.

Et de même la fonction $\Lambda_s(\psi, M, G)$ est mesurable-ϕ.

L'ensemble J des points où $\psi(E)$ a une dérivée Λ relative à la famille G est mesurable. Car d'une part $\Lambda_s - \Lambda_i$ est mesurable, et d'autre part J est l'ensemble où $\Lambda_s - \Lambda_i = 0$. Sur l'ensemble J mesurable-ϕ la dérivée Λ est identique à l'une et à l'autre des fonctions Λ_s et Λ_i, toutes deux mesurables. Donc Λ, définie uniquement sur J, est mesurable-ϕ.

24. THÉORÈME. *Si la fonction d'ensemble $\psi(E)$ est définie et bornée dans le champ $\Delta = \Delta(G)$, l'ensemble I des points de $\Delta(G)$ où l'un des nombres dérivés extrêmes de $\psi(E)$ relativement à la famille régulière G d'ensembles γ est infini, I a la mesure-ϕ nulle.*

Par hypothèse, il existe un nombre positif k tel que $-k < \psi(E) < k$, quel que soit E mesurable-ϕ et inclus dans $\Delta(G)$.

Soit H l'ensemble $\Lambda_s(\psi, M, G) = +\infty$; H est mesurable-ϕ. Supposons $\phi(H) > 0$ et montrons-en l'impossibilité. Soit $G_s(A)$ la famille des γ vérifiant $\psi(\Delta \cdot \gamma) > A\phi(\gamma)$. Elle est régulière comme l'est G. L'ensemble $\Delta[G_s(A)]$, ou par abréviation $\Delta_s(A)$, contient H quel que soit A. Il existe une réunion dénombrable Γ_A d'ensembles γ_i de $G_s(A)$ disjoints, tels que

$$\Delta_s(A) - \Delta_s(A) \cdot \Gamma_A = \Theta_A \text{ a la mesure-}\phi \text{ nulle.}$$

Si $L = \Delta - \Delta \cdot \Gamma_A$, $\psi(\Delta) \geq \psi(\Delta \cdot \Gamma_A) - k$; $\psi(\Delta \cdot \gamma_i) > A\phi(\gamma_i)$ est positif quel que soit i. Les γ_i étant disjoints $\sum_{i<n} \psi(\Delta \cdot \gamma_i)$ est inférieur à k quel que soit n. Donc la série $\Sigma\psi(\Delta \cdot \gamma_i)$ converge et vaut $\psi(\Delta \cdot \Gamma_A)$. En conséquence $\psi(\Delta \cdot \Gamma_A) > A\Sigma\phi(\gamma_i) = A\phi(\Gamma_A)$ et $\psi(\Delta) > A\phi(\Gamma_A) - k$.

Or, Γ_A contient $\Delta_s(A)$ sauf un ensemble de mesure-ϕ nulle. Donc $\phi(\Gamma_A) \geq \phi[\Delta_s(A)]$; $\Delta_s(A)$ contient H. Donc $\phi(\Gamma_A) \geq \phi(H)$ et $\psi(\Delta) > A\phi(H) - k$; $\phi(H)$ étant positif et A un nombre positif quelconque, nous aboutissons à une impossibilité, puisque $\psi(\Delta) < k$.

On démontre de même que l'ensemble $\Lambda_i(\psi, M, G) = -\infty$ a la mesure-ϕ nulle.

25. THÉORÈME. *Si la fonction $\psi(E)$ est définie dans le champ $\Delta(G)$ et si l'égalité $\phi(E) = 0$ entraîne $\psi(E) = 0$, il n'est pas possible qu'il existe deux nombres réels distincts A et B $(A < B)$ tels qu'en tout point M de $\Delta = \Delta(G)$, $\Lambda_i(\psi, M, G) < A < B < \Lambda_s(\psi, M, G)$.*

On considère les deux familles $G_i(A)$, $G_s(B)$ formées des ensembles γ de G vérifiant respectivement $\psi(\Delta\cdot\gamma) < A\phi(\gamma)$ et $\psi(\Delta\cdot\gamma) > B\phi(\gamma)$. Par hypothèse $\Delta[G_i(A)] = \Delta[G_s(B)] = \Delta(G) = \Delta$.

En recouvrant successivement Δ, sauf un ensemble de mesure-ϕ, nulle (et sur lequel par hypothèse $\psi = 0$) avec deux collections dénombrables d'ensembles γ appartenant respectivement à $G_i(A)$ et à $G_s(B)$, on met en évidence les deux inégalités contradictoires $\psi(\Delta) \leq A\phi(\Delta)$ et $\psi(\Delta) \geq B\phi(\Delta)$.

Nous dirons que la fonction d'ensemble $\psi(E)$, définie et bornée dans le champ $\Delta(G)$ y est *métriquement continue*, si l'ensemble E variant indifféramment dans $\Delta(G)$, la condition $\lim \phi(E) = 0$ entraîne $\lim \psi(E) = 0$.

26. Nous dirons qu'au sens de la métrique-ϕ, la famille régulière G d'ensembles γ est *parfaitement régulière* si, quel que soit l'ensemble H mesurable-ϕ inclus dans $\Delta(G)$, la famille G' des ensembles γ de G joints à H a pour ensemble de recouvrement indéfini $\Delta(G')$ l'ensemble H accru au plus d'un ensemble de mesure-ϕ nulle.

THÉORÈME. *Si, la famille G d'ensembles γ étant parfaitement régulière, la fonction d'ensemble $\psi(E)$ est définie, bornée et métriquement continue dans le champ $\Delta = \Delta(G)$, $\psi(E)$ possède une dérivée finie par rapport à la famille G, sur une plénitude-ϕ de Δ.*

Sinon l'ensemble $\Lambda_s(\psi, G) - \Lambda_i(\psi, G) > 0$ aurait une mesure-ϕ positive. On trouverait alors deux nombres A et B, avec $A < B$, (par exemple $A = p/q$, $B = (p+1)/q$, q entier positif étant assez grand et p ayant une certaine valeur entière) de façon que l'ensemble H où simultanément $\Lambda_i < A$, $\Lambda_s > B$, eût une mesure-ϕ positive. Adoptons cette hypothèse. Nous allons prouver $\psi(H) < A\phi(H)$ et pareillement $\psi(H) > B\phi(H)$. La contradiction sera établie.

Soit G' la famille des ensembles γ de G joints à H et $\Delta' = \Delta(G')$ l'ensemble indéfiniment couvert par la famille G' (au sens le la métrique-ϕ). La famille G étant parfaitement régulière, $\Delta' = H + Z$, avec $\phi(Z) = 0$. La fonction $\psi(E)$ s'annulant sous la condition suffisante $\phi(E) = 0$, on trouve, avec $\phi(H) = \phi(\Delta')$, $\psi(H) = \psi(\Delta')$.

Dans la famille G' retenons uniquement des γ (ils sont joints à H) où $\psi(\Delta\cdot\gamma) < A\phi(\gamma)$. Ces γ forment une famille $G'_i(A)$ qui couvre indéfiniment un ensemble $\Delta_A{}^1 = \Delta[G'_i(A)]$ inclus dans Δ' et contenant H. Donc l'ensemble $Z_1 = \Delta_A{}^1 - H$ vérifie $Z_1 \subset Z$ et $\phi(Z) = 0$. Par suite $\phi(H) = \phi(\Delta') = \phi(\Delta_A{}^1)$ et $\psi(H) = \psi(\Delta') = \psi(\Delta_A{}^1)$.

Couvrons $\Delta_A{}^1$, à un ensemble de mesure nulle près, avec une collection

Γ^1 d'ensembles γ_i^1 disjoints appartenant à $G'_i(A)$. Nous pouvons supposer $\phi(\Gamma^1) < \phi(\Delta_A{}^1) + \epsilon = \phi(\Delta') + \epsilon$, avec $\phi[\Gamma_1 \cdot \Delta_A{}^1] = \phi(\Delta_1{}^A) = \phi(\Delta')$, et $\psi(\Gamma^1 \cdot \Delta_A{}^1) = \psi(\Delta_A{}^1) = \psi(\Delta')$. D'après $\psi(\Delta \cdot \gamma_i^1) < A\phi(\gamma_i^1)$, et $\psi(\Delta \cdot \Gamma^1)$ étant fini, la série $\Sigma\psi(\Delta \cdot \gamma_i^1)$ converge et a pour somme $\psi(\Delta \cdot \Gamma^1)$. Donc $\psi(\Delta \cdot \Gamma^1) = \Sigma\psi(\Delta \cdot \gamma_i^1) < A\Sigma\phi(\gamma_i^1) = A\phi(\Gamma^1) < A[\phi(\Delta') + \epsilon]$.

Soit $X^1 = \Gamma^1 - \Gamma^1 \cdot \Delta_A{}^1$; $\Delta_A{}^1$ et X^1 sont disjoints. Donc

$$\phi(X^1) = \phi(\Gamma^1) - \phi(\Delta') < \epsilon.$$

D'après $\Delta_A{}^1 \subset \Delta' \subset \Delta$ et

$$\Gamma_1 = \Gamma_1 \cdot \Delta_A{}^1 + X^1, \quad \Delta \cdot \Gamma^1 = \Delta_A{}^1 \cdot \Gamma^1 + X^1 \cdot \Delta,$$

les deux ensembles $\Delta_A{}^1$ et X^1 étant disjoints:

$$\psi(\Delta \cdot \Gamma^1) = \psi(\Delta_A{}^1 \cdot \Gamma^1) + \psi(X^1 \cdot \Delta),$$

et puisque

$$\psi(\Gamma^1 \cdot \Delta_A{}^1) = \psi(\Delta'), \ \psi(\Delta \cdot \Gamma^1) = \psi(\Delta') + \psi(X^1 \cdot \Delta) < A[\phi(\Delta') + \epsilon].$$

En vertu de l'hypothèse que $\psi(E)$ tend uniformément vers 0 quand $\phi(E)$ tend vers 0, variant dans Δ, $\psi(X^1 \cdot \Delta)$ tend vers 0 avec ϵ, puisqu'il en est ainsi de $\phi(X^1 \cdot \Delta) \leqq \phi(X^1) < \epsilon$. A la limite:

$$\psi(\Delta') \leq A\phi(\Delta').$$

Le raisonnement semblable appliqué à la famille $G'_s(B)$ des ensembles γ de G joints à H (donc inclus dans G') et vérifiant $\psi(\Delta \cdot \gamma) > B\phi(\gamma)$ nous conduit à la conclusion $\psi(\Delta') \geq B\phi(\Delta')$, ce qui est contradictoire avec le précédent résultat si $\phi(\Delta') > 0$. Or $\phi(\Delta') = \phi(H)$. Donc $\phi(H) > 0$ est impossible. La dérivée $\Lambda(\psi, G)$ existe sur une plénitude-ϕ de $\Delta = \Delta(G)$, ensemble indéfiniment couvert, au sens de la métrique-ϕ, par la famille parfaitement régulière G.

Pour l'espace cartésien à r dimensions U_r et la mesure euclidienne borélienne d'ordre r, les familles citées plus haut (18) d'ensembles γ surpassant en mesure une fraction fixe de la sphère minimum les contenant, sont parfaitement régulières.

On voit que l'extension donnée au théorème de Vitali, en rapportant toutes les hypothèse à la seule notion d'une fonction métrique borélienne non négative des ensembles, permet d'énoncer, pour tous les espaces où une telle mesure est définissable, et sans aucun postulat de distance entre points ni de topologie, les analogues de certaines propositions fondamentales relatives à la mesure euclidienne dans les espaces cartésiens ou à la dérivation des fonctions d'une variable réelle.

Observations complémentaires.

La correction des épreuves de cet article me donne l'occasion d'ajouter certaines rectifications et quelque développements à l'étude précédente. Le tout a été d'ailleurs signalé dans une note aux Comptes Rendus de l'Académie des Sciences de Paris (13 novembre 1950, t. 231, p. 1013-1015).

Il s'agit des familles régulières G d'ensembles γ. Nous modifions la définition donnée ci-dessus (26) des familles parfaitement régulières et qui n'a pas d'application.

27. Au sujet du caractère de régularité des familles G d'ensembles γ notons un premier point de détail concernant la quatrième condition. Nous exigeons que la mesure *extérieure-ϕ* de l'ensemble $D = \Sigma\gamma$ soit *finie*, mais non pas que D soit mesurable-ϕ.

Etant donné la famille G d'ensembles γ et si $D = D(G) = \Sigma\gamma$ est non mesurable-ϕ, on peut toujours, en retranchant à certains, sinon à chacun, des γ un ensemble de mesure-ϕ nulle, ramener D à être mesurable-ϕ.

En effet, soit $D = D' + Z$, D' étant mesurable-ϕ et vérifiant $\phi(D') = \phi_i(D)$. Donc $\phi_i(Z) = 0$, et $\phi_e(Z) = \phi_e(D - D') > 0$. ($D'$ et Z sont disjoints).

Les propositions suivantes sont évidentes:

L'ensemble obtenu en retranchant d'un noyau un ensemble de mesure-ϕ nulle est encore un noyau. L'ensemble obtenu en ajoutant à une enveloppe un ensemble de mesure-ϕ nulle est encore une enveloppe.

Tout ensemble inclus dans $\Delta(G)$ et de mesure-ϕ nulle est une enveloppe (comme l'est l'ensemble vide). Toute plénitude-ϕ de $\Delta(G)$ est un ensemble-noyau (comme l'est Δ).

γ étant inclus dans D, $\gamma = \gamma \cdot D = \gamma \cdot D' + \gamma \cdot Z$. Selon un raisonnement déjà rencontré (16), nous voyons que, γ et D', donc aussi $\gamma \cdot D'$, étant mesurables-ϕ, $\gamma \cdot Z$ l'est aussi, et $\phi(\gamma \cdot Z) \leq \phi_i(Z) = 0$.

Soit $\gamma' = \gamma - \gamma \cdot Z$; γ' inclus dans γ est mesurable-ϕ et $\phi(\gamma') = \phi(\gamma)$. A toute collection dénombrable $\Gamma = \Sigma\gamma_i$ d'ensembles γ_i disjoints correspond la collection $\Gamma' = \Sigma\gamma'_i$, incluse dans Γ avec $\phi(\Gamma - \Gamma') = 0$. Les deux familles Γ et Γ' couvrent le même ensemble, à un ensemble de mesure-ϕ nulle près. Or $D' = \Sigma\gamma'$ est mesurable ϕ, avec $\phi(D') = \phi_i(D)$.

Néanmoins il est indispensable de conserver à la quatrième condition toute sa généralité. Car, avec une réunion $D = D(G)$ mesurables-ϕ, on peut avoir des familles partielles G' de G, par exemple formées des γ de G joints à un ensemble H inclus dans $\Delta(G)$ et mesurable-ϕ, de façon que la réunion $D(G')$ des γ de G' soit non-mesurable-ϕ.

28. Donnons de ce cas un exemple emprunté à l'espace linéaire U_1 formé par l'axe Ox, avec la métrique borélienne des longueurs. Soient e_0 un ensemble non mesurable situé sur l'intervalle σ_0 $(0, 1)$ (15 en note), e_1 sur $\sigma_1(1, 2)$, e_2 sur $\sigma_2(2, 3)$ les ensembles respectifs $x + 1$, $x + 2$, si x décrit e_0. Les segments $\alpha\beta$ inclus dans σ_0, définis par les couples de nombres (α, β), peuvent être mis en correspondance par une liaison (A) avec les points de σ_0 chacun à chacun.

Considérons une famille G d'ensembles γ se répartissant comme il suit en trois familles partielles G_1, G_0, G'_0:

1° G_1 est formé des segments s_1 inclus dans σ_1;

2° G_0 est composé des segments s_0 inclus dans σ_0 et correspondant d'après (A) à un point x de $\sigma_0 - e_0$;

3° les s'_0 étant les segments inclus dans σ_0 et correspondant par (A) aux points x de e_0, les ensembles γ de G'_0 sont les segments s'_0 respectivement accrus des deux points $x + 1$ et $x + 2$.

La famille G est régulière. L'ensemble $D = D(G) = \Sigma\gamma$ est $\sigma_0 + \sigma_1 + e_2$, soit l'intervalle $(0, 2)$ diminué du point 1 et accru de l'ensemble e_2 sur σ_2; D est non mesurable. Avec les notations utilisées plus haut, D se décompose en $D' = \sigma_0 + \sigma_1$ et $Z = e_2$; γ' est $\gamma - \gamma \cdot e_2$; $\gamma \cdot e_2$ n'existe pas si γ est dans $G_1 + G_0$ et alors $\gamma' = \gamma$. Si γ est dans G'_0, $\gamma \cdot e_2$ est le point $x + 2$, s'_0 correspondant au point x de e_0; γ' se déduit donc, dans ce cas, de γ par extraction d'un seul point; γ' est la réunion de s'_0 et du point $x + 1$ de e_1. Soient G''_0 la famille des $\gamma' = s'_0 + [x + 1]$ et $G' = G_1 + G_0 + G''_0$; $D(G')$ est $\sigma_0 + \sigma_1$; $D(G')$ est mesurable. Et $\Delta(G') = \Delta(G) = \sigma_0 + \sigma_1$.

Ainsi, il nous a suffi de retrancher un point à certains des ensembles γ composant G pour obtenir une famille G' d'ensembles γ' dont la réunion $D(G')$ est mesurable, tandis que la réunion $D(G)$ des γ formant G ne l'était pas.

Mais il est facile de trouver dans $\Delta(G')$ un ensemble H de longueur déterminée et tel que si $G(H)$ est la famille des γ' de G' joints à H, la réunion $D[G(H)]$ de ces γ' soit un ensemble non mesurable. Il nous suffit de prendre pour H l'intervalle σ_0. Alors $G(H) = G_0 + G''_0$; $D[G(H)] = \sigma_0 + e_1$ est non mesurable.

En conséquence, pour que toute famille partielle d'une famille régulière soit elle-même régulière, il faut garder au quatrième caractère toute sa généralité.

29. La définition des familles parfaitement régulières donnée dans le texte de l'article (26) est trop restrictive pour être utile.

Prenons un exemple. Dans l'espace cartésien U_r à r dimensions, et avec la mesure euclidienne borélienne d'ordre r, soit G une famille d'ensembles γ consistant en intervalles réguliers $(a_i < x_i < b_i + l;\ i = 1, 2, \cdots, r;$ l indépendant de i pour chaque intervalle $\gamma)$ contenus dans une sphère donnée $D = D(G) = \Delta(G)$. Soit H un ensemble mesurable quelconque situé dans D. Soit $G(H)$ l'ensemble des γ de G joints à H. Si H est fermé, $\Delta[G(H)]$, ensemble des points indéfiniment couverts par les γ de $G(H)$, est identique à H. Mais si H n'est pas fermé, $G(H)$ et $G(\bar{H})$ sont identiques. Si donc la mesure de H est inférieure à celle de son ensemble de fermeture $\bar{H}$, $\Delta[G(H)] - H$ a une mesure positive. Par exemple si H est un ensemble dénombrable partout dense dans G, $\Delta[G(H)] = D$, avec mes $H = 0$.

Avec la définition du texte (26), ne seraient parfaitement régulières que les familles G pour lesquelles $\Delta(G)$ aurait la mesure-ϕ nulle. Aussi renonçons-nous à cette première définition et, lui substituant une nouvelle, nous pourrons conserver le théorème du n° 26, dont la démonstration restera entièrement utilisable. Enfin, tous les ensembles mesurables-ϕ contenus dans $\Delta(G)$ vérifieront le théorème fondamental sur l'épaisseur (22).

30. Ne perdons pas de vue que la fonction métrique $\phi(E)$ est de nature très générale. Ce pourra être la mesure euclidienne d'ordre p de Carathéodory dans l'espace cartésien U_r à r dimensions, r surpassant p.

Par exemple, dans l'espace U_3, prenons pour $\phi(E)$ l'aire de l'ensemble E. Une droite a une aire nulle. L'axe des x peut être enfermé sur le plan des xy dans l'ensemble, ouvert dans ce plan: $|y| < \alpha/(1 + x^2)$; α aussi petit qu'on le veut, est proportionnel à l'aire de l'ensemble. Dans U_3 on déformera celui-ci quasi-hélicoïdalement, par une sorte d'enroulement autour de l'axe des x demeurant invariable. Nous obtenons un ensemble $\tilde{\omega}$ relatif à Ox. Dans U_3 les droites parallèles à l'un des trois axes de coordonnées et dont les deux coordonnées constantes sont rationnelles ont une aire totale nulle. Ces droites sont en infinité dénombrable. Si on les énumère, on peut inclure la n^e dans un ensemble $\bar{\omega}_n$ du type décrit ci-dessus pour l'axe des x, et d'aire inférieure à ϵ_n, la série $\Sigma\epsilon_n$ étant convergente. Tel est un ensemble γ constituant de G. On envisagera une infinité non dénombrable de tels ensembles. Pour la régularité de la famille G, la première condition, la mesurabilité de γ [avec $\phi(\gamma) > 0$], est vérifiée. Il faudra s'assurer des trois autres caractères. Mais ces perspectives ouvertes sur le cas le plus simple des espaces cartésiens et des mesures euclidiennes doivent rester présentes à l'esprit si l'on veut saisir toute la portée des notions et des énoncés rencontrés ci-après.

31. Quelle que soit la famille régulière (18) G formée d'ensembles γ, nous distinguons dans $\Delta(G)$ deux classes particulières (non nécessairement disjointes) d'ensembles mesurables-ϕ, à savoir les *ensembles-noyaux* et les *ensembles-enveloppes.*

Selon une notation déjà utilisée, si l'ensemble H est inclus dans $\Delta(G)$, nous désignons par $G(H)$ l'ensemble des γ de G joints à H. Au lieu de $\Delta[G(H)]$ nous écrirons $\delta(H)$. C'est donc l'ensemble indéfiniment couvert au sens de la métrique-ϕ par les γ de G joints à H. Evidement, H étant dans $\Delta(G)$, $\delta(H)$ qui est mesurable-ϕ (14), contient H; $\delta(H) \supset H$. Nous designerons par $\sigma(H)$ l'ensemble $\delta(H) - H$. Donce $\delta(H) = H + \sigma(H)$, H et $\sigma(H)$ étant disjoints.

Nous dirons que, relativement à la métrique-ϕ et à la famille régulière G d'ensembles γ, un ensemble F mesurable-ϕ inclus dans $\Delta(G)$, est un *noyau* quand $\sigma(F) = \delta(F) - F$ a la mesure-ϕ nulle.

Quel que soit l'ensemble γ de G, l'ensemble $\gamma \cdot \Delta$ est un noyau. En effet, soit $H = \gamma \cdot \Delta$. Je dis que $\sigma(H) \subset \rho(\gamma)$, $\rho(\gamma)$ étant l'ensemble défini au n° 18.

Soit $d(\gamma)$ l'ensemble indéfiniment couvert (toujours sous-entendu: au sens de la métrique-ϕ) par les γ' de G' joints à γ; $d(\gamma)$ se compose de deux parties disjointes, la première, $d(\gamma) \cdot \gamma$, située dans γ, la seconde $\rho(\gamma)$ étrangère à γ; $d(\gamma)$ étant dans Δ (comme tout ensemble indéfiniment couvert par des ensembles de G), la première partie de $d(\gamma)$ est commune à γ et à Δ. Elle est dans $\gamma \cdot \Delta$. D'ailleurs tout point M de $\gamma \cdot \Delta$ est dans $d(\gamma)$, puisque, étant dans Δ, il est indéfiniment couvert par des ensembles de G, et que ces ensembles contenant M sont joints à γ. Donc $d(\gamma) \cdot \gamma = \gamma \cdot \Delta = H$ et $d(\gamma) = H + \rho(\gamma)$, H et $\rho(\gamma)$ étant disjoints.

Considérons les γ'' de G composant $G(H)$. Etant joints à H qui est dans γ, ils sont joints à γ. Donc l'ensemble $\delta(H)$ indéfiniment couvert par les γ'' est inclus dans $d(\gamma)$ et par suite $\delta(H) = H + \sigma(H) \subset H + \rho(\gamma)$; H et $\sigma(H)$ étant disjoints, comme le sont d'ailleurs H et $\rho(\gamma)$, on en conclut $\sigma(H) \subset \rho(\gamma)$. Donc $\phi[\sigma(H)] = 0$, d'après le second caractère $\phi[\rho(\gamma)] = 0$ des familles G régulières (18). Donc H est un noyau.

Tout point M de $\Delta(G)$ est un noyau.

Observons d'abord que, la mesure-ϕ étant additive et non négative, $\phi(M)$ existe et vaut zéro. En effet, quel que soit l'ensemble γ de G contenant M, $\phi_e(M) \leq \phi(\gamma)$. Et, M étant dans $\Delta(G)$, la plus petite limite des nombres $\phi(\gamma)$ correspondants est zéro. Donc $\phi_e(M) \leq 0$, soit $\phi(M) = 0$.

La famille $G(M)$ est formée des γ de G contenant M. Pour chacun

d'eux $M \subset \gamma \cdot \Delta$; $\delta(M) = M + \sigma(M)$, ensemble indéfiniment couvert par les γ de $G(M)$, est inclus dans $\delta(\gamma \cdot \Delta) = \gamma \cdot \Delta + \rho(\gamma)$. Donc

$$\phi[\delta(M)] = \phi[\sigma(M)] \leq \phi(\gamma \cdot \Delta) + \phi[\rho(\gamma)] \leq \phi(\gamma).$$

Cette relation étant vérifiée quel que soit γ contenant M, et M étant dans $\Delta(G)$, $\phi[\delta(M)] = 0$; M est un ensemble-noyau.

A deux ensembles-noyaux disjoints F et F' il ne correspond pas nécessairement un nombre positif α tel qu'aucun γ de G, de mesure $\phi(\gamma) < \alpha$, ne puisse être en même temps joint à F et à F'.

Les ensembles $F \cdot \sigma(F')$ et $F' \cdot \sigma(F)$ peuvent ne pas être vides tous les deux. En ce cas α n'existe certainement pas. Si ces deux ensembles sont vides, mais si $\sigma(F) \cdot \sigma(F')$ existe, il peut se faire qu'une suite d'ensembles γ, de mesure $\phi(\gamma)$ tendant vers 0, soient tous joints à F et à F' simultanément.

L'hypothèse que deux ensembles noyaux soient disjoints n'implique rien qui fasse songer à une sorte de distance mutuelle positive des deux ensembles.

32. *La somme d'un nombre fini d'ensembles-noyaux est un ensemble-noyau* (relativement à une famille régulière G).

Soient $F_1, \cdots, F_p$ les ensembles-noyaux donnés et $F = \Sigma F_i$; $G(F) = \Sigma G(F_i)$ (tout γ joint à F est joint à l'un des F_i, ce serait vrai même si les F_i n'étaient pas en nombre fini). $\delta(F) = \Sigma\delta(F_i)$; car un point M indéfiniment couvert par la réunion des $G(F_i)$ est indéfiniment couvert par l'une au moins de ces familles, parce qu'elles sont en nombre fini. Donc $\delta(F) = \Sigma F_i + \Sigma\sigma(F_i) = F + \sigma(F)$, donc $\sigma(F) \subset \Sigma\sigma(F_i)$ et $\phi[\sigma(F)] = 0$.

Si l'ensemble commun à une famille dénombrable d'ensembles-noyaux existe, c'est un ensemble-noyau (relativement à la famille G et à la métrique-ϕ).

Soit $H_1, H_2, \cdots, H_n, \cdots$ une suite d'ensembles-noyaux relativement à G et $H = \Pi H_n$ leur ensemble commun majeur. Les H_n sont mesurables-ϕ. La métrique étant borélienne, H est mesurable-ϕ, s'il n'est pas vide. Si $H_n = H + J_n$, H et J_n étant disjoints, l'ensemble commun aux J_n est vide, puisque H est l'ensemble majeur commun aux H_n. Formons $G(H)$. C'est l'ensemble des γ de G joints à H. Un tel γ est *a fortiori* joint à H_n, quel que soit n, puisque H_n contient H. Donc $G(H) \subset G(H_n)$ et $\delta(H) \subset \delta(H_n)$, soit :

$$H + \sigma(H) \subset H_n + \sigma(H_n) = H + J_n + \sigma(H_n).$$

H et $\sigma(H)$ sont disjoints, et il en est de même de H, J_n et $\sigma(H_n)$. Donc $\sigma(H) \subset J_n + \sigma(H_n)$.

$\sigma(H)$ est donc commun à tous les $J_n + \sigma(H_n)$. Un point M de $\sigma(H)$ ne peut pas être commun à tous les J_n, puisque ΠJ_n est vide. Donc M est commun à tous les $\sigma(H_n)$; $\sigma(H) \subset \sigma(H_n)$. Or, H étant un noyau, $\phi[\sigma(H_n)] = 0$. Donc $\phi[\sigma(H)] = 0$. H est relativement à G un ensemble-noyau.

En particulier si les H_n décroissent ($H_n \supset H_{n+1}$) et si $\phi(H_n)$ qui est non croissant a pour limite le nombre λ, ou bien $\lambda > 0$ et alors le noyau $H = \Pi H_n$ existe et a pour mesure λ; ou bien $\lambda = 0$ et $H = \Pi H_n$ peut ne pas exister. S'il existe, sa mesure-ϕ est 0.

33. La condition que la famille des ensembles-noyaux soit dénombrable est essentielle. Par exemple, dans l'espace linéaire U_1 de l'axe des x, et avec la métrique borélienne des longueurs, soit e_0 un ensemble situé sur l'intervalle $s_0(0, 1)$ et de longueur extérieure 1, de longueur intérieure 0; e_0 a la puissance du continu, et il en est de même de $\sigma \cdot e_0$, quel que soit le segment σ contenu dans s_0; cet ensemble $\sigma \cdot e_0$ a la longueur extérieure σ et la longueur intérieure 0.

Considérons une famille G d'ensembles γ, G se décomposant en familles $G_{k,h}$ ($k \geq 1$, $h = 1, \cdots, 2^k$) ainsi constituées: Tout ensemble γ de $G_{k,h}$ est le segment $\sigma_{k,h}$ $[(h-1)2^{-k} \leq x \leq h2^{-k})$ diminué d'un point de $e_0 \cdot \sigma_{k,h}$. Pour chaque système d'indices (k, h), les γ de $G_{k,h}$ sont en infinité ayant la puissance du continu. La famille $G = \sum_{k,h} G_{k,h}$ est composée d'ensembles γ ayant chacun une longueur déterminée. La famille G est régulière. Chacun des γ de G est un noyau.

$\Delta(G)$, ensemble indéfiniment couvert au sens des longueurs par les γ de G, est identique au segment $(0, 1)$. Si l'on considère la famille des γ de $G_{k,h}$, famille non dénombrable, composée d'ensembles-noyaux relativement à G, l'ensemble commun à tous les ensembles de cette famille est $\sigma_{k,h} - \sigma_{k,h} \cdot e_0$, qui est non mesurable, puisque $\sigma_{k,h} \cdot e_0$ est non mesurable. Cet ensemble commun n'étant pas mesurable, n'est pas un ensemble-noyau.

34. Relativement à la métrique-ϕ et à une famille régulière G d'ensembles γ, nous appelons *ensemble-enveloppe le complémeniaire par rapport à* $\Delta = \Delta(G)$ *d'un ensemble-noyau.*

C'est là une première définition. En voici une seconde, équivalente à la première:

Relativement à la métrique-ϕ *et à la famille régulière* G *d'ensembles* γ, *l'ensemble* O *mesurable-*ϕ, *inclus dans* Δ, *est un ensemble-enveloppe si la*

famille $G'(O)$ des γ de G disjoints de $\Delta - O$ est telle que les points de O indéfiniment couverts par la famille $G - G'(O) = G(\Delta - O)$, complémentaire de $G'(O)$ par rapport à G, forment un ensemble $\tau(O)$ de mesure-ϕ nulle.

Un ensemble O' satisfaisant à la première définition, donc complémentaire d'un ensemble noyau $F = \Delta - O'$ satisfait à la seconde.

En effet, la famille $G - G'(O') = G(\Delta - O')$ est $G(F)$. Posons $\tau(O') = \sigma(F)$. Les points de $O' = \Delta - F$ converts indéfiniment par $G(F)$ forment bien l'ensemble $\tau(O')$ et $\phi[\tau(O')] = \phi[\sigma(F)] = 0$.

Réciproquement, soit O vérifiant la seconde définition et $F' = \Delta - O$. Je dis que F' est noyau. En effet, $G(\Delta - O)$ est $G(F')$, et puisque par hypothèse l'ensemble $\tau(O)$ est celui des points $O = \Delta - F'$ couverts indéfiniment par $G(F')$, $\sigma(F')$ est identique à $\tau(O)$ et d'après l'hypothèse $\phi[\tau(O)] = 0$, on conclut $\phi[\sigma(F')] = 0$; F' est noyau.

Voici une remarque utilisée dans ma définition initiale des ensembles-enveloppes, donnée dans ma Note aux *Comptes Rendus* du 13 novembre 1950:

En vertu de notre seconde définition de l'ensemble-enveloppe O, si M est un point de $O - \tau(O)$, les γ de G contenant M et étrangers à $G'(O)$ ont leurs mesures $\phi(\gamma)$ bornées inférieurement par un nombre positif α, dépendant évidemment de M.

Car les γ de G étrangers à $G'(O)$ sont dans $G(\Delta - O)$; M étant dans $O - \tau(O)$ n'est pas indéfiniment couvert par la famille $G(\Delta - O)$, en sorte que les γ de $G(\Delta - O)$ contenant M vérifient bien la condition énoncée.

Un point de $\tau(O)$ est indéfiniment couvert par $G(\Delta - O)$. Il l'est ou ne l'est pas par $G'(O)$.

35. Corollaires. 1°. *L'ensemble commun à un nombre fini d'ensembles enveloppes est un ensemble-enveloppe.*

2°. *La somme d'une infinité dénombrable d'ensembles-enveloppes est un ensemble-enveloppe.*

Si $O_1, O_2, \cdots, O_n, \cdots$ sont des enveloppes, les ensembles $F_n = \Delta - O_n$ sont des noyaux. $\Pi O_n = \Delta - \Sigma F_n$, $\Sigma O_n = \Delta - \Pi F_n$.

Si les O_n sont en nombre fini, ΣF_n est un ensemble-noyau; ΠO_n est une enveloppe. Que les O_n soient en nombre fini ou en infinité dénombrable, ΠF_n est un noyau, donc ΣO_n est une enveloppe.

L'exemple du n° 33 montre que, si les ensembles enveloppes étaient en infinité indénombrable, leur somme ne serait plus nécessairement une enveloppe.

3°. *Si F est un ensemble-noyau et O un ensemble enveloppe, $F - F \cdot O$ est noyau, $O - O \cdot F$ est enveloppe.*

Car $F - F \cdot O = F \cdot (\Delta - O)$ est l'intersection de deux noyaux; $O - O \cdot F = O \cdot (\Delta - F)$ est l'intersection de deux enveloppes.

4°. *Si F ensemble-noyau est inclus dans O, ensemble-enveloppe, l'ensemble Θ indéfiniment couvert par les γ de G joints à F et disjoints de $\Delta - O = f$ vérifie les inégalités (égalités non exclues)* :

$$F - F \cdot \tau(O) \subset \Theta \subset F + O \cdot \sigma(F).$$

La famille des γ joints à F [donc inclus dans $G(F)$] et disjoints de f [donc inclus dans $G'(O)$] est $G(F) \cdot G'(O) = g$.

Un point M de F étranger à $\tau(O)$ est dans $O - \tau(O)$; M est donc couvert indéfiniment par des ensembles γ de $G'(O)$. Ces γ contenant M sont dans $G(F)$. Ces γ sont dans g. Donc M est dans Θ et $\Theta \supset F - F \cdot \tau(O)$.

D'autre part, tout point de Θ doit être couvert indéfiniment par $G(F)$, donc il est dans $F + \sigma(F)$. Il n'est pas dans f, dont tous les γ de $G'(O)$, et *a fortiori* de g, sont disjoints; donc il est dans $O \cdot [F + \sigma(F)] = F + O \cdot \sigma(F)$. La seconde inclusion est établie.

$\Delta = \Delta(G)$ *est à la fois noyau et enveloppe par rapport à la famille régulière* G. Car $G(\Delta) = G'(\Delta) = G$; $\sigma(\Delta)$ et $\tau(\Delta)$ sont vides.

Dans le plan U_2, la fonction métrique $\phi(E)$ étant l'aire borélienne, et si la famille G est composée des carrés contenus dans un ensemble ouvert D, les noyaux F sont les ensembles fermés diminués d'un ensemble d'aire nulle (ou inexistant), les enveloppes sont les ensembles ouverts accrus d'un ensemble d'aire nulle.

36. *Nature antitopologique des espèces considérées.* Il est à remarquer qu'à l'inverse de l'ordre adopté en topologie commune, la notion d'ensemble-noyau, évoquant, mais très vaguement, celle d'ensemble fermé des espaces cartésiens, se définit directement, tandis que celle d'ensemble-enveloppe, qui rappellerait, de loin, celle d'ensemble ouvert, s'introduit par la voie de la complémentarité et se rattache à la première. En topologie générale, le caractère de l'ensemble fermé est intrinsèque. Il se conserve dans une extension de l'espace contenant l'ensemble, pourvu que la topologie reçoive l'extension correspondante. Au contraire, l'ensemble ouvert de l'espace initial cesse d'être ouvert dans l'extension. Quel que soit l'espace variable contenant un ensemble fermé immuable, le complémentaire de celui-ci par rapport à l'espace variable est toujours un ensemble ouvert, mais dont la nature ne cesse pas de changer avec l'espace. Au contraire l'ensemble fermé garde constamment son caractère.

Une topologie qui donnerait la première place aux caractères intrinsèques des ensembles ne se fonderait pas sur les ensembles ouverts.

37. L'ensemble $\Delta = \Delta(G)$ est défini comme étant l'ensemble des points de l'espace U indéfiniment couverts par les ensembles γ de la famille G incluse dans U. Mais un point M de Δ n'est caractérisé que par la famille $G(M)$ des γ de G contenant ce point M, $G(M)$ couvrant indéfiniment M. Aussi cette famille $G(M)$ caractérise-t-elle, solidairement avec M, la totalité des autres points de $\Delta[G(M)] = \delta(M) = M + \sigma(M)$, à savoir la totalité des points N de $\sigma(M)$ qui, distincts de M, sont indéfiniment couverts, avec M, par $G(M)$.

Ecrivons $N \sim M$, cela signifiant que N est dans $\sigma(M)$. Mais, puisque parmi les γ contenant M, il en existe une infinité contenant N, donc appartenant à $G(N)$, et dont la mesure $\phi(\gamma)$ tend vers zéro, il en résulte que M est dans $\sigma(N)$. Donc la relation $N \sim M$ entraîne $M \sim N$. Cette liaison, exprimée par le signe $\sim$, est réflexive, selon le langage usuel d'aujourd'hui. Nous disons que M et N sont *indiscernables* dans Δ et relativement à la famille G.

Parmi les points de $\sigma(M)$ distincts de N, il peut s'en trouver un, soit P, tel que les γ de $G(M)$ contenant P ne couvrent pas indéfiniment N, tandis qu'ils couvrent indéfiniment P, puisque P est par hypothèse dans $\sigma(M)$. En ce cas, il pourra se faire que $G(P)$ ne couvre pas indéfiniment N, ni corrélativement que $G(N)$ ne couvre pas indéfiniment P, $G(N)$ et $G(P)$ couvrant indéfiniment l'un et l'autre le point M. Alors les deux couples de relations réversibles $(N \sim M, M \sim N)$ et $(P \sim M, M \sim P)$ n'entraînent pas $(P \sim N, N \sim P)$. Tout cela est évidemment très éloigné des conditions d'une topologie.

Nous dirons que M et N sont *totalement indiscernables par* G si, ces deux points étant dans $\Delta = \Delta(G)$, les deux ensembles $\delta(M)$ et $\delta(N)$ sont identiques et si de plus les γ de $G(M)$ ne contenant pas N et les γ de $G(N)$ ne contenant pas M ont leurs mesures $\phi(\gamma)$ bornées inférieurement par un nombre positif: $\delta(M) = \delta(N) = \Delta[G(M) \cdot G(N)]$.

38. Relativement à la métrique-ϕ dans l'espace U et à la famille régulière G d'ensembles γ, appelons *clos métrique* d'un ensemble H contenu dans $\Delta = \Delta(G)$ l'ensemble $\delta(H) = \Delta[G(H)]$ indéfiniment couvert par les γ de G joints à H. L'ensemble $\delta(H)$ qui contient H, joue par rapport à H un role qui fait songer à celui de l'ensemble de fermeture $\bar{E}$ d'un ensemble E

en topologie. Mais nous trouvons une opposition capitale entre les deux notions dans le fait que $\bar{\bar{E}} = \bar{E}$ (un ensemble de fermeture est identique à son propre ensemble de fermeture), tandis qu'il n'y a aucune raison logique pour que $\delta[\delta(H)]$, le clos métrique de $\delta(H)$, l'ensemble indéfiniment couvert par les γ de G joints au clos métrique $\delta(H)$ de H, soit identique à $\delta(H)$.

Notons que, d'après $\delta(H) = H + \sigma(H)$,

$$\delta[\delta(H)] = \delta(H) + \delta[\sigma(H)] = \delta(H) + \sigma(H) + \sigma[\sigma(H)] = \delta(H) + \sigma[\sigma(H)].$$

Le clos métrique $\delta(F)$ d'un noyau F a même mesure que celui-ci. Mais on ne voit pas que ce clos métrique soit nécessairement un noyau. De ce que $\phi[\sigma(F)] = 0$, il ne paraît y avoir aucune nécessité pour que la mesure-ϕ de $\sigma[\sigma(F)]$ soit elle aussi nulle. S'il n'en est pas ainsi, $\delta(F)$ n'est pas un noyau, bien que F le soit.

39. A la lumière de l'étude précédente, revenons aux propositions des n^{os} 22 et 28.

Soit F un ensemble noyau de la famille régulière G formée d'ensembles γ. La totalité des γ de G joints à F constitue la famille $G(F)$ couvrant indéfiniment $\delta(F) = F + \sigma(F)$, F et $\sigma(F)$ étant disjoints avec $\phi[\sigma(F)] = 0$.

Soit M un point de F. Les épaisseurs diverses (supérieure, inférieure, éventuellement exacte) de F au point M (22) relatives à la famille G sont les limites du rapport $\cdot \phi(F \cdot \gamma)/\phi(\gamma)$, l'ensemble γ de G contenant M et $\phi(\gamma)$ tendant vers 0. Mais, puisque G ne renferme aucun ensemble γ étranger à $G(M)$ et contenant M, il y a identité entre les épaisseurs de F en son point M relativement à G et relativement à $G(F)$. Le théorème 22 sur l'épaisseur de l'ensemble $\Delta = \Delta(G)$ relativement à G est vrai de l'épaisseur de l'ensemble $\delta(F)$ relativement à $G(F)$, puisque $\delta(F) = \Delta[G(F)]$. En conséquence, l'épaisseur de $\delta(F)$ relativement à $G(F)$ est 1 sur une plénitude-ϕ de $\delta(F)$. Soit H cette plénitude de $\delta(F)$; H est inclus dans $\delta(F)$ et $\phi(H) = \phi[\delta(F)] = \phi(F)$, puisque $\phi[\sigma(F)] = 0$, F étant noyau.

D'après $H \subset F + \sigma(F)$, $H = H \cdot F + H \cdot \sigma(F)$, et, avec $\phi[H \cdot \sigma] = 0$, $\phi(H \cdot F) = \phi(H) = \phi(F)$; $H' = H \cdot F$ est une plénitude-ϕ de F. En tout point M de H', l'épaisseur de $\delta(F) = F + \sigma(F)$ relative à $G(F)$ est 1.

D'autre part, pour tout ensemble γ, $\gamma \cdot \delta(F) = \gamma \cdot F + \gamma \cdot \sigma(F)$. Donc $\gamma \cdot \delta(F)$ a même mesure que $\gamma \cdot F$; donc l'épaisseur de $\delta(F)$ et celles de F relativement à $G(F)$ sont égales en tout point. En conséquence, l'épaisseur de F par rapport à $G(F)$ est 1 en tout point de H'. Mais, comme nous l'avons vu, l'épaisseur de F par rapport à $G(F)$ ou par rapport à G est la même en tout point de F. Donc, *relativement à la famille G, l'épaisseur de*

l'ensemble-noyau F est 1 *sur une plénitude-ϕ de F. Je dis que, sur une plénitude de l'ensemble-enveloppe $O = \Delta - F$, l'épaisseur de F est* 0.

En effet, $\tau(O)$ étant $\sigma(F)$, soit M un point de $O - \tau(O) = \Delta - \delta(F)$; $O - \tau(O)$ a même mesure que O; M étranger à $\delta(F)$ n'est pas couvert indéfiniment par $G(F)$. Les γ de G contenant M et joints à F ont (33) leur mesure $\phi(\gamma)$ bornée inférieurement par un nombre positif α, dépendant de M et de F). Si γ contient M et si $\phi(\gamma) < \alpha$, $\phi(F \cdot \gamma) = 0$, puisque $F \cdot \gamma$ est vide. Donc $\lim \phi(\gamma \cdot F)/\phi(\gamma) = 0$ quand γ contenant M a sa mesure $\phi(\gamma)$ tendant vers 0. L'épaisseur de F en tout point M de $O - \tau(O)$ est nulle. Or d'après $\phi[\tau(O)] = \phi[\sigma(F)] = 0$, $O - \tau(O)$ est une plénitude-ϕ de $O = \Delta - F$.

40. Au n° 26 nous avons considéré un ensemble H tel que, avec les notations de la presente étude complémentaire, l'ensemble $\Delta[G(H)] = \delta(H) = H + \sigma(H)$ ait même mesure que H. Un tel ensemble est donc, dans le cas de toute famille régulière G, un ensemble-noyau. Sans autre hypothèse sur G que sa régularité, en-dehors des points isolés de Δ, des ensembles $\gamma \cdot \Delta$, et des sommes finies de ces variétés, nous ne sommes assurés de l'existence que d'un seul ensemble-noyau relativement à G, à savoir $\Delta = \Delta(G)$ lui-même.

Nous avons montré, et il n'y a rien à retrancher à la démonstration, que, si la fonction d'ensemble $\psi(E)$ est définie, bornée, et métriquement continue dans le champ $\Delta(G)$ (25), les hypothèses indépendantes $\Lambda_i(\psi, M, G) < A$, $\Lambda_s(\psi, M, G) > B$ vérifiées en tout point M d'un ensemble noyau H (A et B ne dépendent pas de M), entraînent respectivement $\psi(H) < A\phi(H)$, $\psi(H) > B\phi(H)$.

$\Lambda_i(\psi, M, G)$, $\Lambda_s(\psi, M, G)$ sont, au point M, les nombres dérivés inférieur et supérieur, relativement à la famille G, de la fonction d'ensemble $\psi(E)$ (23). On verrait comme plus haut qu'en un point M de H, les nombres dérivés par rapport à G et par rapport à $G(H)$ sont les mêmes. Notons que, d'après $\phi[\sigma(H)] = 0$ et la fonction d'ensemble $\psi(E)$ étant métriquement continue, quel que soit γ dans G: $\psi[\gamma \cdot \delta(H)] = \psi(\gamma \cdot H)$, (comme dans le cas particulier où la fonction ψ est la mesure ϕ).

41. Enfin, venons -en à la définition rectifiée des familles G d'ensembles γ de l'espace U, parfaitement régulières relativement à la métrique-ϕ adoptée dans U.

La famille G sera dite *parfaitement régulière*, si, ϵ étant un nombre positif indépendant, tout ensemble H mesurable-ϕ inclus dans Δ contient un

ensemble-noyau F et est inclus dans un ensemble-enveloppe O, de façon que $\phi(H) - \phi(F)$ et $\phi(O) - \phi(H)$ soient inférieurs à ϵ.

Tout ensemble-noyau F étant mesurable-ϕ, F est contenu dans un ensemble-enveloppe O vérifiant $\phi(O) < \phi(F) + \epsilon$. Et tout ensemble enveloppe O' de mesure-ϕ supérieure à ϵ contient un ensemble-noyau F' avec $\phi(F') > \phi(O') - \epsilon$.

Ces propriétés ne sont nullement certaines avec des familles G simplement, mais non parfaitement, régulières.

Quand la famille G est parfaitement régulière, il est possible, relativement à tout ensemble H mesurable-ϕ, inclus dans Δ, de trouver une famille $g(H)$ d'ensembles γ de G, telle que l'ensemble $\omega(H) = \Delta[g(H)]$, formé des points indéfiniment couverts par les γ de $g(H)$ ne diffère de H que par un ensemble de mesure-ϕ nulle.

La suite numérique ϵ_n tendant vers 0, soient F_n noyau, O_n enveloppe, deux ensembles vérifiant $F_n \subset H \subset O_n$, $\phi(O_n) - \phi(F_n) < \epsilon_n$. D'après les théorèmes du n° 32 et 35, on peut supposer F_n croissant (non décroissant) et O_n décroissant (non croissant). Posons $f_n = \Delta - O_n$; f_n est noyau et croît avec n. Soit g_n la famille $G(F_n) \cdot G'(O_n)$ et Θ_n l'ensemble $\Delta(g_n)$. Nous avons vu (35) que

$$F_n - F_n \cdot \tau(O_n) \subset \Theta_n \subset F_n + O_n \cdot \sigma(F_n).$$

Nous posons $g(H) = \Sigma g_n$.

Désignons par ρ_p la famille $\sum_{n>p} g_n$; $g(H) = g_1 + g_2 + \cdots + g_p + \rho_p$. Pour $n > p$, $G'(O_n)$ disjoint de f_n est *a fortiori* disjoint de $f_p \subset f_n$. *A fortiori* g_n est-il disjoint de f_p, donc inclus dans $G'(O_p)$. En conséquence, $\Delta(\rho_p) \subset O_p$.

Soit $\omega_n = \Theta_1 + \Theta_2 + \cdots + \Theta_n$.

$\omega(H) = \omega_p + \Delta(\rho_p)$. Donc, d'après $F_1 + F_2 + \cdots + F_p = F_p$ et $F_p + O_p = O_p$,

$$\omega(H) \supset F_p - \sum_{n \leq p} F_n \cdot \tau(O_n) \quad \text{et} \quad \omega(H) \subset \sum_{n \leq p} O_n \cdot \sigma(F_n) + O_p.$$

Quand p croît, F_p tend en croissant vers $\sum_{n \geq 1} F_n = H - \lambda$, avec $\phi(\lambda) = 0$; O_p tend en décroissant vers $\prod_{n \geq 1} O_p = H + \mu$, avec $\phi(\mu) = 0$.

Donc

$$H - \lambda - \sum_{n \geq 1} F_n \cdot \tau(O_n) \subset \omega(H) \subset H + \mu + \sum_{n \geq 1} O_n \cdot \sigma(F_n).$$

Les $\phi(\tau)$ et $\phi(\sigma)$ étant nuls. $H - H \cdot \omega(H)$ et $\omega(H) - H \cdot \omega(H)$ ont l'un et l'autre la mesure-ϕ nulle, $\omega(H)$ étant $\Delta[g(H)]$.

42. THÉORÈME. *Relativement à une famille G parfaitement régulière, l'épaisseur de tout ensemble H mesurable-ϕ inclus dans $\Delta(G)$ est 1 sur une plénitude-ϕ de H et 0 sur une plénitude-ϕ de $\Delta - H$.*

Car, quel que soit l'ensemble F inclus dans H, l'épaisseur de F en un quelconque de ses points est au plus égale à l'épaisseur de H au même point. Si F est un ensemble-noyau, son épaisseur relativement à G est 1 sur une plénitude-ϕ de F. Or $\phi(F)$ peut être supposé plus grand que $\phi(H) - \epsilon$. Donc, l'ensemble $J(H)$ des points de H où l'épaisseur $\eta(H, M)$ est 1, a une mesure supérieure à $\phi(H) - \epsilon$ quel que soit $\epsilon > 0$. Donc, $J(H)$ a pour mesure $\phi(H)$. C'est une plénitude-ϕ de H.

$\Delta - H$ étant mesurable-ϕ a l'épaisseur 1 sur une plénitude-ϕ de lui même. Donc, sur une plénitude de $\Delta - H$, H a l'épaisseur 0. Car évidemment $\eta_i(H) + \eta_s(\Delta - H) = 1$, $\eta_s(H) + \eta_i(\Delta - H) = 1$, en tout point M de Δ.

43. THÉORÈME. *Relativement à la famille G parfaitement régulière selon une métrique-ϕ, toute fonction d'ensemble $\psi(E)$ définie, bornée, métriquement continue dans le champ $\Delta = \Delta(G)$ a une dérivée finie sur une plénitude-ϕ de Δ.*

Sans quoi, il existerait (26) deux nombres A, B avec $A < B$ et un ensemble H de mesure-ϕ positive tel qu'en tout point M de H, $\Lambda_i(\psi, M, G) < A$, $\Lambda_s(\psi, M, G) > B$. Si H n'est pas un noyau relativement à G, et si G est simplement, mais non parfaitement régulière, nous ne pouvons rien tirer de ces conclusions. Car $\phi[\sigma(H)] > 0$. Par contre, G étant parfaitement régulière et, d'après $\phi(H) > 0$, H contient un ensemble-noyau F de mesure-ϕ positive. Or, les deux hypothèses précédentes, vérifiées l'une et l'autre sur le noyau F sont incompatibles, puisqu'elles entraîneraient $\psi(F) < A\phi(F)$ et $\psi(F) > B\phi(F)$.

OBSERVATION. *Si, la famille G étant parfaitement régulière, en tout point M d'un ensemble H mesurable-ϕ inclus dans $\Delta = \Delta(G)$ la fonction d'ensemble $\psi(E)$, définie, bornée et métriquement continue dans le champ Δ vérifie la condition $\Lambda_i(\psi, M, G) < A$* [ou la condition $\Lambda_s(\psi, M, G) > B$], *A (ou B) étant un nombre indépendant de M variant sur H, on en conclut $\psi(H) < A\phi(H)$* [ou $\psi(H) > B\phi(H)$].

Traitons l'hypothèse $\Lambda_i < A$ sur H.

H étant mesurable-ϕ et inclus dans Δ contient un ensemble-noyau F et

si $J = H - F$, on peut supposer $\phi(J) < \epsilon$. L'inégalité $\Lambda_i < A$ est vérifiée sur F; F étant noyau, on en conclut (40) $\psi(F) < A\phi(F)$. Donc

$$\psi(H) = \psi(F) + \psi(J) < A[\phi(H) - \phi(J)] + \psi(J).$$

On peut faire tendre $\phi(J)$ vers 0, par une succession de choix de F; $\psi(E)$ étant métriquement continue, $\psi(J)$ tend vers 0 avec $\phi(J)$. Donc $\psi(H) \leq A\phi(H)$. D'ailleurs le signe d'égalité peut être supprimé. Car si, en tout point de H, $\Lambda_i < A$, il existe un α positif tel que l'ensemble des points de H où $\Lambda_i < A - \alpha$ a pour mesure-ϕ un nombre positif λ, sans quoi $\Lambda_i \geq A$ sur une plénitude-ϕ de H. Donc

$$\psi(H) \leq A[\phi(H) - \lambda] + (A - \alpha)\lambda < A\phi(H).$$

Pareillement l'hypothèse $\Lambda_s > B$ en tout point de H entraîne

$$\psi(H) > B\phi(H).$$

44. THÉORÈME. *Si, relativement à une métrique-ϕ et à une famille G d'ensembles γ parfaitement régulière, la fonction d'ensemble $\psi(E)$ est définie, bornée et métriquement continue dans le champ $\Delta = \Delta(G)$, la fonction $\Lambda(M)$ égale à la dérivée $\Lambda(\psi, M, G)$ aux points de Δ où cette dérivée existe et est finie, et à zéro sur l'ensemble complémentaire R, $\Lambda(M)$ est sommable-ϕ sur Λ et, quel que soit E mesurable-ϕ inclus dans Δ, $\psi(E)$ est égal à $\mathfrak{I}_L[\Lambda(M)d\phi, E]$, intégrale lebesguienne de $\Lambda(M)d\phi$ sur E.*

La démonstration est du type le plus classique. Puisque $\psi(E)$ a relativement à G une dérivée finie sur une plénitude-ϕ de Δ, $\phi(R) = 0$ et en conséquence $\psi(R) = 0$.

Soient E^+ et E^- les ensembles des points de E où respectivement $\Lambda(M) > 0$ et $\Lambda(M) < 0$; $\psi(E) = \psi(E^+) + \psi(E^-)$. Car, sur $E' = E - E^+ - E^-$ (E' contient $E \cdot R$), $\Lambda(M) = 0$ et par suite $\psi(E') = 0$, indépendamment de la mesure-ϕ de E' [Sur E', $-\epsilon < \Lambda(M) < \epsilon$ quel que soit $\epsilon > 0$ indépendant, donc $-\epsilon\phi(E') < \psi(E') < \epsilon\phi(E')$].

Divisons l'intervalle $(0, +\infty)$ par une suite de nombres croissant indéfiniment l_i ($i = 0, 1, 2, \cdots$) avec $l_0 = 0$, $l_i - l_{i-1} < \omega$, ω étant indépendant de i (≥ 1). Soit E_i l'ensemble des points de E (de E^+) où $l_{i-1} < \Lambda(M) \leq l_i$; $E^+ = \Sigma E_i$ ($i \geq 1$) et $l_{i-1}\phi(E_i) < \psi(E_i) \leq l_i\phi(E_i)$ (43), donc $\psi(E_i) = \lambda_i\phi(E_i)$, λ_i étant un certain nombre vérifiant $l_{i-1} < \lambda_i \leq l_i$. Les $\psi(E_i)$ étant positifs, et la fonction $\psi(E)$ étant bornée dans le champ Δ, la série $\Sigma\psi(E_i)$ converge et $\psi(E^+) = \Sigma\psi(E_i) = \Sigma\lambda_i\phi(E_i)$. Donc $\Lambda(M)$ est sommable-ϕ sur E^+ et $\psi(E^+) = \mathfrak{I}_L[\Lambda(M)d\phi, E^+]$.

De même $\Lambda(M)$ est sommable-ϕ sur E^- et $\psi(E^-) = \mathfrak{J}_L[\Lambda(M)d\phi, E^+]$. $\Lambda(M)$ étant sommable-ϕ sur E^+, sur E^- et aussi sur E' où $\Lambda(M) = 0$, $\Lambda(M)$ est sommable-ϕ sur E et $\psi(E) = \mathfrak{J}_L[\Lambda(M)d\phi, E]$.

E peut être identique à Δ. Donc $\Lambda(M)$ est sommable-ϕ sur Δ.

45. THÉORÈME (réciproque ou précédent). *Si, la famille G étant parfaitement régulière relativement à la métrique-ϕ, la fonction de point $f(M)$ est définie et sommable-ϕ sur l'ensemble $\Delta = \Delta(G)$, l'intégrale lebesguienne $\mathfrak{J}_L[f(M)d\phi, E] = \psi(E)$, où E est un ensemble quelconque mesurable-ϕ, inclus dans Δ, est une fonction d'ensemble définie, bornée et métriquement continue dans le champ Δ, et la dérivée de $\psi(E)$ relativement à G est $f(M)$ sur une plénitude-ϕ de Δ.*

On suppose la fonction $f(M)$ finie en tout point de Δ.

La fonction d'ensemble $\psi(E)$ est définie dans le champ Δ, au sens donné à cette expression (23); $\psi(E)$ est bornée. Car, Δ^+ étant l'ensemble $f > 0$ et Δ^- l'ensemble $f < 0$,

$$|\psi(E)| < \mathfrak{J}_L[|f(M)|\, d\phi, \Delta] = \mathfrak{J}_L[f(M)d\phi, \Delta^+] - \mathfrak{J}_L[f(M)d\phi, \Delta^-] < \infty.$$

Enfin $\psi(E)$ est métriquement continue dans Δ. Car, si $\phi(E)$ tend vers 0, $\psi(E) = \mathfrak{J}_L[f(M)d\phi, E]$ tend vers 0. Car f étant sommable-ϕ, cette dernière intégrale tend vers 0.

En conséquence, relativement à G et sur un ensemble $\Delta - R$ qui est une plénitude-ϕ de Δ, $\psi(E)$ a une dérivée finie. $\Lambda(M)$ étant cette dérivée sur $\Delta - R$ et 0 sur R, nous venons de montrer que $\Lambda(M)$ est sommable-ϕ sur Δ et que $\psi(E) = \mathfrak{J}_L[\Lambda(M)d\phi, E]$ quel que soit E mesurable-ϕ dans Δ. Donce si $g(M) = \Lambda(M) - f(M)$, $\mathfrak{J}_L[g(M)d\phi, E] = 0$ quel que soit E mesurable-ϕ avec $E \subset \Delta$. En conséquence, si petit que soit $\epsilon > 0$, les ensembles $g(M) > \epsilon$ et $g(M) < -\epsilon$ ont la mesure-ϕ nulle. Donc $g(M) = 0$ et $f(M) = \Lambda(\psi, M, G)$ sur une plénitude-ϕ de Δ.

Il ressort de tout ce qui précède que les propriétés des fonctions absolument continues d'une variable réelle s'étendent aux fonctions d'ensembles métriquement continues relatives aux familles parfaitement régulières, dans un espace et avec une métrique borélienne entièrement quelconques.

PARIS, FRANCE.

ON NORMAL OPERATORS IN HILBERT SPACE.*

By C. R. PUTNAM.

---

1. For definitions and terminology used in this paper, see [3], especially pp. 27 ff.

The following Lemma is a slight generalization of a result obtained recently by Fuglede [1] (cf. also [2]) concerning normal operators:

LEMMA. *Let N_1 and N_2 be two normal operators in a Hilbert space with spectral resolutions*

$$N_j = \int z\, dK_z^j \qquad (z = x + iy;\ j = 1, 2). \tag{1}$$

If A is any bounded operator satisfying

$$AN_1 \subseteqq N_2 A, \tag{2}$$

then

$$AK_z^1 = K_z^2 A, \text{ for all } z, \tag{3}$$

and hence,

$$AN_1^* \subseteqq N_2^* A. \tag{4}$$

If $N_1 = N_2 = N$, the Lemma becomes exactly the Theorem of Fuglede.

The Lemma will be instrumental in proving a theorem concerning the equivalence of bounded normal operators in Hilbert space. The theorem in question is the following:

THEOREM I. *Let N_1 and N_2 denote two bounded normal operators for which there exists a non-singular bounded operator T such that*

$$N_2 = TN_1T^{-1}. \tag{5}$$

Then N_1 and N_2 are unitarily equivalent, that is, there exists a unitary operator U such that

$$N_2 = UN_1U^*. \tag{6}$$

In case both operators N_1 and N_2 are Hermitian or both are unitary, Theorem I is known. For the Hermitian case, the result is apparently due to Toeplitz, while the Theorem, in case both N_1 and N_2 are unitary, is due

---

* Received September 23, 1950.

to Wintner [7], p. 149. He has pointed out that a difficulty arises if one attempts to transcribe the proofs for these known cases to the present case, where N_1 and N_2 are supposed only normal ([8], pp. 139-140), and that the Theorem is surely true if N_1 and N_2 are finite normal matrices ([7], p. 149).

2. *Proof of the Lemma.* The proof of the Lemma will be given by indicating the appropriate modifications of the argument used by Fuglede [1]. Let α denote an arbitrary Borel set in the complex plane and define the operators $K^1(\alpha)$ and $K^2(\alpha)$ by

$$K^j(\alpha) = \int_\alpha dK_z^j, \qquad j = 1, 2.$$

Suppose now that α and β are disjoint squares of the lattice considered in [1] and let f denote an arbitrary element of $\mathfrak{D}_{N_1}$. Since A is bounded, the set $\mathfrak{D}_{N_1}$ is contained in the set $\mathfrak{D}_{AN_1}$, and, since $K^1(\alpha)f$ belongs to $\mathfrak{D}_{N_1}$, relation (2) implies

$$AN_1K^1(\alpha)f = N_2AK^1(\alpha)f. \tag{7}$$

An application of the bounded operator $K^2(\beta)$ to each side of (7) and a utilization of the spectral resolutions (1) show that

$$K^2(\beta)A\int_\alpha z\, dK_z^1 f = \int_\beta z\, dK_z^2\, AK^1(\alpha)f, \tag{8}$$

for all f in $\mathfrak{D}_{N_1}$. An argument parallel to that used in [1] shows that (8) implies

$$K^2(\beta)AK^1(\alpha) = 0; \tag{9}$$

the Lemma follows easily from (9), cf. [1], and the proof is complete.

3. *Proof of Theorem* I. If λ denotes an arbitrary complex number and if I is the identity operator, then (5) holds if and only if $N_2 + \lambda I = T(N_1 + \lambda I)T^{-1}$. Since $N_j + \lambda I$ ($j = 1$ and 2) are normal, it is clear that it may be assumed that N_1 and N_2 are non-singular. This assumption will be convenient later in the proof.

For $j = 1, 2$, let $N_j = P_jU_j$ denote the polar decomposition of N_j, so that P_j is Hermitian positive-definite, U_j is unitary, and $P_jU_j = U_jP_j$. Equation (5) can be expressed as

$$TN_1 = N_2T \quad \text{or} \quad TP_1U_1 = P_2U_2T, \tag{10}$$

and consequently, by the Lemma,

$$TN_1^* = N_2^*T \quad \text{or} \quad TU_1^*P_1 = U_2^*P_2T. \tag{11}$$

The following set of equations is an obvious consequence of (10) and (11):

(12) $$TP_1^2 = TP_1U_1U_1^*P_1 = P_2U_2TU_1^*P_1 = P_2U_2U_2^*P_2T = P_2^2T.$$

Hence,

(13) $$TP_1 = P_2T;$$

this equation follows from (12) and from the fact that P_j, as the unique positive definite square root of P_j^2, can be expressed as a power series in P_j^2. (Relation (13) can also be obtained as a consequence of the Lemma.) Hence, by (10) and (13), $P_2U_2T = TP_1U_1 = P_2TU_1$, and therefore, since P_2 is non-singular,

(14) $$TU_1 = U_2T$$

If $T = PU$ denotes the polar decomposition of T, it follows from (13) and (14) that $UP_1 = P_2U$ and $UU_1 = U_2U$. In the latter (unitary) case, this result is given in [7], p. 149; the proof in the Hermitian case is similar. In fact, (13) implies

(15) $$PUP_1 = P_2PU,$$

and consequently, on taking the transposed conjugate of each side of (15), $P_1U^*P = U^*PP_2$. Multiplication on the left of each side of the last equation by PU and the use of (15) yield $P_2P^2 = P^2P_2$ and hence, as above, $P_2P = PP_2$. It follows from (15), that $PUP_1 = PP_2U$ and therefore $UP_1 = P_2U$, which was to be shown. Hence,

$$N_2 = P_2U_2 = UP_1U^*UU_1U^* = UP_1U_1U^* = UN_1U^*$$

that is, (6). This completes the proof of Theorem I.

4. The following theorem will be derived as a consequence of Theorem III (see below):

THEOREM II. *Let A denote an arbitrary bounded matrix and let B denote a bounded normal matrix. If* (i) $AB - BA = C$ *and if* (ii) $CB = BC$, *then* $C = 0$.

Even without the assumption (ii), the relation g. l. b. $| x^*Cx | = 0$, where x denotes a vector on the Hilbert sphere $\| x \| = 1$, is valid; cf. [4]. If, in addition, (ii) is assumed, it is clear that Theorem II greatly refines the assertion of the preceding sentence.

That the conclusion of Theorem II is false if B is not required to be normal is easily seen by verifying that the matrices

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

satisfy (i) and (ii), although $C \neq 0$.

The following theorem will be proved in Section 5, by using Fuglede's theorem together with an adaptation of his methods:

THEOREM III. *In a Hilbert space, let B be a normal operator, with a spectral resolution*

$$B = \int z\, dK_z, \tag{16}$$

and let A, C be arbitrary operators satisfying

$$AB \subseteq BA + C, \tag{17}$$

(i) *If, in addiaion to* (17), *C satisfies*

$$C \text{ is bounded and } \neq 0 \tag{18}$$

and

$$CB \subseteq BC, \tag{19}$$

then A is unbounded. Furthermore, B cannot have an eigenfunction f belonging to $\mathfrak{D}_A$ and satisfying $Cf \neq 0$.

(ii) *If, in addition to* (17), *C satisfies $C = \lambda I + \mu B \neq 0$, where I is the identity operator and λ, μ denote complex constants, then A is unbounded. Furthermore, B cannot have an eigenfunction which belongs to $\mathfrak{D}_A$ and is associated with an eigenvalue z_0 satisfying $\lambda + \mu z_0 \neq 0$.*

Remark. If f is an eigenfunction of B belonging to $\mathfrak{D}_A$, it is clear that $Bf \subset \mathfrak{D}_A$ and hence, by (17), $f \subset \mathfrak{D}_C$.

Theorem II is an obvious consequence of part (i) of Theorem III.

It is known that if

$$AB = BA + \lambda I, \qquad \lambda \neq 0, \tag{20}$$

then A and B cannot both be bounded Hermitian operators; [9] or [5]. It follows from (ii) of Theorem III above that this result can be strengthened to: if A and B are self-adjoint ([3], p. 34) and satisfy (20) in the operator sense, then neither A nor B can be bounded. This is clear from the fact that the validity of (20) in the operator sense implies $AB \subseteq BA + \lambda I$ and $BA \subseteq AB - \lambda I$. In addition, it follows from (ii) above, since the number μ occurring there is zero, that neither A nor B can possess an eigenfunction belonging to the domain of the other. Rellich [5] has obtained results, which are rather explicit as far as the nature of the operators A and B is concerned, under different assumptions; namely, where (20) holds on a certain dense set and where A and B are merely symmetric (as distinguished from the more restrictive requirement of being self-adjoint) but are subject to certain other conditions.

Remark. It may be mentioned that, whether or not A or B is Hermitian, still (20) implies that A and B cannot *both* be bounded; cf. [6] and also the first paragraph of [4], wherein is cited a reference to [9].

5. *Proof of* (i) *of Theorem* III. Suppose, if possible, that A is bounded. For any Borel set α of the complex plane, define $K(\alpha) = \int_\alpha dK_z$. If f is in the domain of B (hence of AB) and if α and β have the same significance as in Section **2**, it follows that

$$K(\beta)A\int_\alpha z dK_z f = \int_\beta z dK_z\, AK(\alpha)f + K(\beta)CK(\alpha)f. \tag{21}$$

Assumption (19) of Theorem III and an application of Fuglede's theorem (specifically the Lemma of Section **1** in which $N_1 = N_2$ and A become identified with the present B and C respectively), imply that $CK(\alpha) = K(\alpha)C$. Hence the second term on the right of (21) becomes $K(\beta)K(\alpha)Cf$. Since α and β are disjoint, it follows from the properties of the projection operators that $K(\beta)K(\alpha) = 0$; so that (8), in which the superscripts are omitted, is valid for all f in the domain of B. As in Section **2**, this leads to (9) and hence to $AK_2 = K_2A$ for all z; cf. (3). The last equality implies however that $AB \subseteq BA$ (cf. [1], p. 38, end of section 2). Since A is bounded, $\mathfrak{D}_{AB} = \mathfrak{D}_B$. Thus, the relations $AB \subseteq BA$ and (17) imply $Cf = 0$ for all f in $\mathfrak{D}_B$, which is dense in the containing Hilbert space, $\mathfrak{H}$. Thus, since C is bounded, $Cf = 0$ for all $f \subset \mathfrak{H}$ $(= \mathfrak{D}_C)$. That is, $C = 0$, in contradiction with assumption (18) of Theorem III. Hence, A is unbounded and the proof of the first statement of Theorem III is complete.

Suppose, if possible, that z_0 is an eigenvalue of B with an eigenfunction f belonging to $\mathfrak{D}_A$ such that $Cf \neq 0$. The symbol $K(z_0)$, where z_0 is a Borel set consisting of one point, has been defined above (cf. also [3], p. 53). An application of the operator equation (17) to f (cf. the Remark following the statement of Theorem III) and a subsequent application of the bounded operator $K(z_0)$ to the resulting functional equation yield

$$K(z_0)ABf = K(z_0)BAf + K(z_0)Cf. \tag{22}$$

In virtue of $K(z_0)C = CK(z_0)$, which follows from the theorem of Fuglede, and of the equations

$$Bf = z_0 f, \qquad K(z_0)B = z_0K(z_0), \qquad K(z_0)f = f,$$

(cf. [3], p. 53), relations (22) implies that

$$z_0K(z_0)Af = z_0K(z_0)Af + Cf, \tag{23}$$

in contradiction to the assumption $Cf \neq 0$. Thus, the supposition at the beginning of this paragraph is untenable and the proof of part (i) of Theorem III is complete.

Proof of (ii) *in Theorem* III. It is clear from the assumption of (ii) that C is normal and hence $K(\alpha)C \subseteq CK(\alpha)$ for an arbitrary Borel set α. By a proof essentially identical with that given in the first paragraph of this section, it follows that A is unbounded. (In the present case, it should be noted that the assumption that A is bounded again leads to the conclusion that $Cf = 0$ for all $f \subset \mathfrak{D}_B$ $(= \mathfrak{D}_C)$. Since C is normal, it is densely defined and closed. Consequently, Cf is defined and is the zero element for *all* f in $\mathfrak{H}$. Thus, C is the (bounded) zero operator in $\mathfrak{H}$, a contradiction.)

In order to prove the second assertion of (ii), suppose, if possible, that the assertion is false. That is, suppose that B has an eigenvalue z_0, where $\lambda + \mu z_0 \neq 0$, with an associated eigenfunction f in $\mathfrak{D}_A$. Proceeding as in the proof of (i) one obtains

$$z_0 K(z_0) Af = z_0 K(z_0) Af + (\lambda + \mu z_0) f, \tag{24}$$

corresponding to (23). Relation (24) leads to a contradiction and the proof of Theorem III is complete.

THE INSTITUTE FOR ADVANCED STUDY.

REFERENCES.

[1] B. Fuglede, "A commutativity theorem for normal operators," *Proceedings of the National Academy of Sciences*, vol. 36 (1950), pp. 35-40.

[2] P. R. Halmos, "Commutativity and spectral properties of normal operators," *Acta Scientiarum Mathematicarum*, vol. 12 (1950), pp. 153-156.

[3] B. v. Sz. Nagy, *Spektraldarstellung linearer transformationen des Hilbertschen Raumes*, Ann Arbor, 1947.

[4] C. R. Putnam, "On commutators of bounded matrices," *American Journal of Mathematics*, vol. 73 (1951), pp. 127-131.

[5] F. Rellich, "Der Eindeutigkeitssatz für die Lösungen der quantenmechanische Vertauschungsrelationen," *Göttinger Nachrichten*, 1946, pp. 107-115.

[6] H. Wielandt, "Ueber die Unbeschränktheit der Operatoren der Quantenmechanik," *Mathematische Annalen*, vol. 121 (1949), p. 21.

[7] A. Wintner, "On non-singular bounded matrices," *American Journal of Mathematics*, vol. 54 (1932), pp. 145-149.

[8] ———, "On normal inertia," *Studia Mathematica*, vol. 8 (1939), pp. 135-140.

[9] ———, "The unboundedness of quantum-mechanical matrices," *Physical Review*, vol. 71 (1947), pp. 738-739.

SOME RETRACTION PROPERTIES OF THE ORBIT DECOMPOSITION SPACES OF PERIODIC MAPS.*

By E. E. Floyd.

Let X be a finite-dimensional locally compact metric space. Let T be a periodic map on X. There is generated a decomposition of X whose elements are the orbits $[T^i(x) \mid i = 0, 1, \cdots]$ of points x in X. There is then generated a corresponding decomposition space X^* called the modular space or the orbit decomposition space of the pair X, T. We show that several connectedness and local connectedness properties are transmitted from X to X^*. As a consequence, we prove that if X is either an absolute neighborhood retract or an absolute retract, then X^* has the same property.

Our results depend in a very fundamental fashion on the theory of P. A. Smith concerning periodic maps [6, 7, 8, 9]. Hence we assume on the part of the reader a general knowledge of his methods and results.

Unless otherwise stated, *we assume that X is a locally compact n-dimensional metric space, $n < \infty$, and that T is a periodic map of X onto X of prime period p.* It will be pointed out in Theorem 5 that the preceding results hold when p is not prime. Denote by X^* the collection of all orbits in X, and by $\Lambda: X \to X^*$ the function which assigns to each x in X the orbit containing x. If we give to X^* the customary decomposition topology, then Λ is a continuous open map of X onto X^*. It may be shown that the map Λ is also compact; i. e., if $C \subset X^*$ is compact, then $\Lambda^{-1}(C)$ is compact. We denote by L the fixed point set of T, and define $L^* = \Lambda(L)$.

If X is a locally compact space, and G an abelian group, then X is *acyclic over G* if every Čech cycle on a compact subset of X over G bounds on a compact subset of X (we shall agree that each coordinate of a 0-cycle has coefficient sum 0). We denote by I the group of integers, and by I_p the group of integers modulo p.

Theorem 1. *If X is acyclic over I_p, then so is X^*.*

Proof. We shall base the homology theory of X on the complete family of special coverings of X [8, p. 136]. All projections will be invariant projections [8, p. 138]. The homology theory for X^* will be based on the coverings of the type $\mathfrak{U}^* = \Lambda\mathfrak{U}$, where $\mathfrak{U}$ is special.

* Received October 9, 1950.

Let W^* be a compact set in X^*. We construct compact sets V^* and U^* with $W^* \subset V^* \subset U^*$, with every cycle in $V^* \cap L^*$ bounding in $U^* \cap L^*$, and with every cycle in W^* bounding in V^* mod $V^* \cap L^*$. As we point out at the end of the proof, this is sufficient to prove that every cycle in W^* bounds in U^*.

Let $W = \Lambda^{-1}(W^*)$. We pick V^* so that every ρ-cycle in W ρ-bounds in $V = \Lambda^{-1}(V^*)$ (for definition of a ρ-cycle, see [7, pp. 361-362]). That this is possible may be seen by examining the proof of a theorem of Smith [6, Theorem II]. Then we pick U^* so that every cycle in $V \cap L$ bounds in $U \cap L$, where $U = \Lambda^{-1}(U^*)$. This is possible since L is acyclic over I_p [6, Theorem II].

The sets U^*, V^*, W^* have the properties of the second paragraph of this proof. It is clear that every cycle in $V^* \cap L^*$ bounds in $U^* \cap L^*$, since Λ is a homomorphism on L. Now let $\gamma = \{\gamma(\mathfrak{U}^*)\}$ be a cycle in W^*. Consider the cycle $\Gamma = \{\Gamma(\mathfrak{U})\} = \{\Lambda^{-1}\gamma(\mathfrak{U}^*)\}$ [8, p. 140]. Following Smith, we let $\delta = 1 - T$ and $\sigma = 1 + T + \cdots + T^{p-1}$. Then Γ is a δ-cycle in W, and hence Γ δ-bounds in V. Moreover, there exist chains $X(\mathfrak{U})$ in W with $\Lambda X(\mathfrak{U}) = \gamma(\mathfrak{U})$ and with $\Gamma(\mathfrak{U}) = \sigma X(\mathfrak{U})$. Since Γ δ-bounds in V, it follows from a theorem of Smith [8, p. 144] that $\gamma \sim 0$ in V^* mod $V^* \cap L^*$. Hence the conditions of the second paragraph are fulfilled.

To prove that every cycle in W^* bounds in U^*, note the following diagram where all homomorphisms are inclusion homomorphisms, and I_p is the coefficient group:

$$\begin{array}{ccccc} & & & & H_n(W^*) \\ & & & \overset{j}{\swarrow} & \\ H_n(V^* \cap L^*) & \xrightarrow{k} & H_n(V^*) & \xrightarrow{i} & H_n(V^*, V^* \cap L^*) \\ \downarrow l & & \downarrow l_1 & & \\ H_n(U^* \cap L^*) & \xrightarrow{k_1} & H_n(U^*) & \to & H_n(U^*, U^* \cap L^*) \end{array}$$

Let $x \in H_n(W^*)$. Then, by construction, $ij(x) = 0$. By exactness, there exists $y \in H_n(V^* \cap L^*)$ with $k(y) = j(x)$. But $l(y) = 0$. Hence $l_1 k(y) = k_1 l(y) = 0$. Then $l_1 j(x) = 0$. We have proved that every cycle in W^* bounds in U^*, and the theorem follows.

As is customary, we say that a locally compact space M is *lc* over the group G if given a neighborhood U of $x \in M$, there exists a neighborhood V of x with $V \subset U$ and with every cycle in V bounding in U.

THEOREM 1′. *If X is lc over I_p, so also is X^*.*

Proof. We indicate the minor changes we make in the proof of the preceding theorem. We need only prove that X^* is *lc* at points $y \in L^*$ since

Λ is a local homomorphism at points of $X - L$. Using the notation of the previous theorem, we have U^* preassigned. Since L is *lc* [9, p. 703], we may pick V^* so that every cycle in $V \cap L$ bounds in $U \cap L$. We then pick W^* so that every ρ-cycle in W ρ-bounds in V. The proof then goes exactly as before.

THEOREM 2. *If X is acyclic over I and over I_p, then so also is X^*.*

Proof. Let W^* be a compact set in X^*. Let $W = \Lambda^{-1}(W^*)$. We pick a compact set V^* with $V^* \supset W^*$ and with every cycle in W^* over I_p bounding in V^*. We may suppose, moreover, that every cycle in W over I bounds in $V = \Lambda^{-1}(V^*)$. Let U^* be a compact set with $U^* \supset V^*$ and with every cycle in V^* over I_p bounding in U^*. We shall prove that every cycle in W^* over I bounds in U^*.

We first note that if z is a cycle over I in W^*, then $pz \sim 0$ in V^*. For $\Lambda^{-1}z$ is a cycle in W with $\Lambda\Lambda^{-1}z = pz$ [8, p. 141]. Since $\Lambda^{-1}z$ bounds in V, then pz bounds in V^*.

If z is a chain over the integers, we denote by $r(z)$ the chain over I_p obtained by replacing each coefficient of z by the corresponding integer mod p. Let K be a complex and z a cycle in K over I. Suppose c is a chain with $\phi c = pz$, where ϕ is the boundary operator. Then $r(c)$ is a cycle over I_p. Moreover if $z \not\sim 0$ then $r(c) \not\sim 0$ [1, p. 220 ff.].

Since every cycle in V^* over I_p bounds in U^*, then if $\mathfrak{U}^*$ is a covering of X^* there exists a covering $\mathfrak{V}^*$ such that if π denotes a projection of $\mathfrak{V}$ into $\mathfrak{U}$ then for any cycle $z(\mathfrak{V}^*)$ over I_p in V^* we have $\pi z(\mathfrak{V}^*) \sim 0$ in U^*. Suppose now that $z = \{z(\mathfrak{U}^*)\}$ is a cycle over I in W^*. We have $pz \sim 0$ in V^*. Let $c(\mathfrak{V}^*)$ be a chain in V^* with $\phi c(\mathfrak{V}^*) = pz(\mathfrak{V}^*)$. Then $\phi\pi c(\mathfrak{V}^*) = p\pi z(\mathfrak{V}^*)$. But $r[\pi c(\mathfrak{V}^*)] = \pi r[c(\mathfrak{V}^*)] \sim 0$ in U^*. Hence $\pi z(\mathfrak{V}^*) \sim 0$ in U^* so that $z(\mathfrak{U}^*) \sim 0$ in U^*. Then $z \sim 0$ in U^* so that the theorem follows.

THEOREM 2′. *If X is lc over I, then so also is X^*.*

Proof. We note that X is also *lc* over I_p [2, p. 11], and hence so also is X^*. The proof is then practically the same as that of Theorem 2, where U^* is now a compact closed neighborhood of a point of L^* and V^* and W^* are then picked so that U^*, V^*, W^* have the properties of the above proof.

THEOREM 3. *Suppose X is arcwise connected, and that $L \neq 0$. Then if X is simply connected, so also is X^*.*

Proof. Let $x \in L$ and let $y = \Lambda(x)$. Let $f: I \to X^*$ be a path in X^*, where I is the unit interval, such that $f(0) = f(1) = y$. Then we consider Λ as being a light interior map of the compact space $\Lambda^{-1}f(I)$ onto $f(I)$. Then there exists a map $g: I \to X$ with $\Lambda g = f$ [3]. Then $g(0) = g(1) = x$, since $\Lambda^{-1}(y) = x$. Then g is a closed path in X and hence $g \backsim 0$ in X. Then $f \backsim 0$ in X^* and the assertion follows.

THEOREM 3'. *If X is locally connected and locally simply connected, then so is X^*.*

Proof. The proof is similar to the above, and may be omitted.

THEOREM 4. *If X is an absolute neighborhood retract* [*an absolute retract*], *then X^* is also an absolute neighborhood retract* [*absolute retract.*]

Proof. Suppose X is an *ANR*. Then X is LC^1 and *lc* over I. By Theorems 2 and 3, X^* is LC^1 and *lc* over I. Then by a theorem of Hurewicz [5] we have that X^* is LC^n and hence an *ANR*. Suppose X is an *AR*. Then X is simply connected and acyclic over I and I_p. By Theorem 2, X^* is acyclic over I. Moreover $L \neq 0$ [6]. Then by Theorem 3, X^* is simply connected. It follows, then, that X^* is an *AR* [4].

We shall now point out that we may, in our results, replace the cyclic transformation group of prime order p generated by T by any finite solvable transformation group. Let G be a finite transformation group on X. Let H be a normal subgroup of G. Let X^*_1 be the orbit decomposition space of the pair X, H and $\Lambda_1: X \to X^*_1$ the corresponding decomposition map. The points of X^*_1 are sets of the form $H(x)$, $x \in X$. If $G_1 = G/H$, if $h = \{g\}$ is an equivalence class in G_1 and if $H(x) \in X^*_1$, define $h[H(x)] = (gH)(x) = H(gx) \in X^*_1$. In this way G_1 acts as a transformation group on X^*_1. Call the resulting orbit decomposition space X^*_2 and denote by $\Lambda_2: X^*_1 \to X^*_2$ the corresponding decomposition map. Finally, let X^* be the orbit decomposition space of the pair X, G and $\Lambda: X \to X^*$ the decomposition map.

LEMMA 1. *There exists a homeomorphism S of X^* onto X^*_2 with $S\Lambda = \Lambda_2\Lambda_1$.*

Proof. Let $x \in X$. Then $\Lambda^{-1}\Lambda(x) = G(x)$. It may be seen, letting $\Gamma = \Lambda_2\Lambda_1$, that $\Gamma^{-1}\Gamma(x) = G(x)$. It follows that the function $\Gamma\Lambda^{-1}$ is a 1-1 function of X^* onto X^*_2. Moreover it is bicontinuous since both Γ and Λ are continuous and open. Setting $S = \Gamma\Lambda^{-1}$, we obtain the conclusion.

THEOREM 5. *Let G be a finite solvable transformation group on X. Let X^* denote the orbit decomposition space of the pair X, G. Then if X is an absolute neighborhood retract [absolute retract], so also is X^*.*

Proof. Let $\Lambda: X \to X^*$ be the decomposition map. According to the previous lemma, we may write $\Lambda = \Lambda_m \cdot \cdot \cdot \Lambda_2 \Lambda_1$ where each Λ_i is equivalent to the decomposition map generated by a prime periodic map. The theorem then follows from Theorem 4.

Let us also note the following problems:

PROBLEM 1. *Does Theorem 5 hold when G is not solvable?*

PROBLEM 2. *If X is a manifold, does X^* contain about each point arbitrarily small contractible neighborhoods? Does, as a matter of fact, X contain arbitrarily small contractible invariant neighborhoods about each of its fixed points?*

PROBLEM 3. *If X is a 3-manifold, is there an arbitrarily small invariant 2-sphere about each fixed point?*

UNIVERSITY OF VIRGINIA.

BIBLIOGRAPHY.

1. P. Alexandroff and H. Hopf, *Topologie*, Berlin (1935).
2. E. Čech, "Sur la connexité locale d'ordre supérieur," *Compositio Mathematica*, vol. 2 (1935), pp. 1-25.
3. E. E. Floyd, "Some characterizations of interior maps," *Annals of Mathematics*, vol. 51 (1950), pp. 571-575.
4. W. Hurewicz, "Beiträge zur Topologie der Deformationen. II. Homotopie- und Homologiegruppen," *Nederlandsche Akademie van Wetenschappen, Proceedings*, vol. 38 (1935), pp. 521-528.
5. W. Hurewicz, "Homotopie, Homologie und lokaler Zusammenhang," *Fundamenta Mathematicae*, vol. 25 (1935), pp. 467-485.
6. P. A. Smith, "Fixed point theorems for periodic transformations," *American Journal of Mathematics*, vol. 63 (1941), pp. 1-8.
7. ———, "Fixed points of periodic transformations," Appendix B in Lefschetz's *Algebraic Topology*, New York (1942).
8. ———, "Transformations of finite period," *Annals of Mathematics*, vol. 39 (1938), pp. 127-164.
9. ———, "Transformations of finite period. II," *Annals of Mathematics*, vol. 40 (1939), pp. 690-711.

ON THE NON-EXISTENCE OF CONJUGATE POINTS.*

By AUREL WINTNER.

1. Let $f(t)$ be a real-valued, continuous function on a t-interval, I, which need not be closed or bounded. Consider only those solutions $x(t)$ of the differential equation

$$x'' + f(t)x = 0 \tag{1}$$

on I which are real-valued and are distinct from the trivial solution ($\equiv 0$). If no solution of (1) has more than one zero on I, then (1) will be called disconjugate on I. This terminiology is suggested by the fact that, in calculus of variations, the disconjugate nature of (1) on I defines the absence of conjugate points on I, if (1) is Jacobi's equation.

The remarks collected in this note will center about the following criterion: (1) is disconjugate on I if and only if there exists on I *some* function $y = y(t)$ possessing a continuous first derivative which satisfies the inequality

$$y'(t) + y^2(t) \leqq -f(t) \tag{2}$$

at every t contained in I.

If the $\leqq$ is refined to an $<$ in (2), the sufficiency of the resulting criterion is contained in a more general result of Bôcher [1], who has applied it in order to derive Sturm's "separation" and "comparison" theorems. But the classical proof of these theorems on (1) is simple enough and, if they are granted, both the necessity and the sufficiency of the criterion (2) follow as immediate corollaries. For the sake of completeness, this will be shown, in a generalized form, in Section **10.**

Sturm's comparison theorem follows by observing that if $f_1(t) \geqq f_2(t)$, and if (2) is satisfied by a y for $f = f_1$, then it is satisfied, by the same y, for $f = f_2$. But it does not seem to have been noticed in the literature that the criterion (2) easily leads to non-trivial results also. For instance, it implies, without the *ad hoc* considerations of the known proofs, and almost as trivialities, the explicit sufficient criteria which go back to Liapounoff (cf. [2]) and Kneser [4], respectively. In addition, other sufficient criteria also follow from the same source.

* Received August 20, 1950.

2. First, it will be shown that the two explicit criteria, just mentioned, are manifestations of one and the same situation, the t-range, I, being (α) a bounded interval in the first case and (β) a half-line in the second. If a^+ denotes $\max(0, a)$, these two criteria can be formulated as follows:

The differential equation (1) must be disconjugate on $I = [0, 1]$ if

$$(\alpha) \qquad \int_0^1 f^+(s)\,ds \leqq 4,$$

and it must be disconjugate on $I = (0, \infty)$ if

$$(\beta) \qquad \int_t^\infty f^+(s)\,ds \leqq 1/(4t) \text{ for } 0 < t < \infty$$

(the convergence of the last integral is part of the assumption, since $f^+ \geqq 0$).

Ad (α). Since $f^+(t) \geqq f(t)$, the differential equation (1) is a Sturmian minorant of the differential equation $x'' + f^+(t)x = 0$. On the other hand, $f^+(t) \geqq 0$. Hence, it is sufficient to show that if (α) is replaced by the conditions

$$(3) \qquad \int_0^1 f(s)\,ds \leqq 4,\ f(t) \geqq 0, \text{ where } 0 \leqq t \leqq 1,$$

then (1) is disconjugate on $I = [0, 1]$. To this end, define for $0 < t < 1$ a function $y(t)$ by placing

$$(4) \qquad y(t) = \int_t^1 f(s)\,ds + \begin{cases} t^{-1} - 4 \text{ if } 0 < t \leqq \frac{1}{2}, \\ (t-1)^{-1} \text{ if } \frac{1}{2} \leqq t < 1. \end{cases}$$

This $y(t)$ has a continuous derivative for $0 < t < 1$. If (3) is assumed, it is readily verified that the function (4) satisfies (2) for $0 < t < 1$. Hence, (3) is sufficient in order that (1) be disconjugate on $I = (0, 1)$. Finally, the transition from this open interval I to the corresponding closed interval, $I = [0, 1]$, can be effected by using the continuity of $f(t)$ on $[0, 1]$ and observing that in (2) the sign of inequality must hold on some subinterval of $(0, 1)$ by virtue of (3) and (4).

Ad (β). Corresponding to the transition from (α) to (3), it can be assumed that (β) is replaced by

$$(4') \qquad \int_t^\infty f(s)\,ds \leqq 1/(4t) \text{ and } f(t) \geqq 0, \text{ where } 0 < t < \infty.$$

Then a direct substitution shows that (2) is satisfied by

$$y(t) = \int_t^\infty f(s)\,ds + 1/(4t). \tag{5}$$

It is clear from these verifications that both (3) and (4′) can be improved by suitable "logarithmic" functions. In addition, (β) can be modified as follows:

(β^*) If $\int^\infty f(t)\,dt$ is convergent (possible just conditionally) and, as a function of its lower limit of integration, satisfies the inequality

$$\left| \int_t^\infty f(s)\,ds \right| \leqq 1/(4t) \tag{4*}$$

on the half-line I, then (1) is disconjugate on I.

In other words, the restriction $f(t) \geqq 0$ can be omitted in (4′). In fact, it is readily seen that the $y(t)$ defined by (5) satisfies (2) by virtue of (4*) alone.

Actually, (4*) can be improved to

$$-\tfrac{3}{4} \leqq t \int_t^\infty f(s)\,ds \leqq \tfrac{1}{4}. \tag{4**}$$

In fact, (2) holds for the function (5) whenever

$$\left| \int_t^\infty f(s)\,ds + 1/(4t) \right| \leqq 1/(4t^2)^{\frac{1}{2}}.$$

But this inequality can be written in the form

$$-1/(2t) \leqq \int_t^\infty f(s)\,ds + 1/(4t) \leqq 1/(2t),$$

which is equivalent to (4**).

It should be mentioned that the constant $-\frac{3}{4}$ occurring in (4**) will arise in Section **9** in a different context.

3. For $I = (0, \infty)$, a new sufficient condition, of a type quite different from (4), is as follows:

$$[\int_t^\infty f(s)ds]^2 \leqq f(t)/4, \text{ where } 0 < t < \infty. \tag{6}$$

The sufficiency of (6) is readily verified by choosing $y(t) = 2\int_t^\infty f(s)ds$ in (2).

The constant factor $\frac{1}{4}$ occurring in (4'), (4*), (6) is the best absolute constant. In fact, if $f(t) = c/t^2$, then (4'), (4*), (6) are satisfied when $4c \leqq 1$, but the $\frac{1}{4}$ must be replaced by a somewhat larger constant if $4c = 1 + \epsilon \to 1 + 0$. Hence it is sufficient to ascertain that the case $f(t) = c/t^2$ of (1) cannot be disconjugate on $(0,\infty)$ if $4c > 1$. But this is well-known (Kneser). In fact, the case $f(t) = c/t^2$ of (1) is satisfied, for any real c and for $0 < t < \infty$, by $x(t) = t^b$, and therefore by the real part of t^b as well, if $b(b-1) + c = 0$. Since this quadratic equation has real roots only if $4c \leqq 1$, it follows that the real solutions $x(t)$ must have zeros clustering at $t = \infty$ if $4c > 1$.

If the half-line $I = (0,\infty)$ is replaced by a bounded interval, say by $I = [0, a]$, then a sufficient condition corresponding to (6) is

$$[\int_0^t f(s)ds]^2 \leqq f(t)/4, \text{ where } 0 \leqq t \leqq a. \tag{7}$$

In fact, (7) implies that (2) is satisfied, for $0 \leqq t \leqq a$, by $y(t) = -2\int_0^t f(s)ds$.

4. Let I be a half-line reaching to ∞, say $I = (t_0,\infty)$. If (1) has a solution the zeros of which cluster at infinity, then (1) is called oscillatory, and otherwise non-oscillatory. It is clear from Sturm's separation theorem that if (1) is oscillatory, then every solution of (1) has zeros clustering at infinity.

For the same reason, it is easy to see that if (1) is non-oscillatory, then there must exist a sufficiently large t^0 $(\geqq t_0)$ having the property that no solution of (1) possesses more than one zero on the half-line (t^0,∞). Consequently, (1) is non-oscillatory if and only if there exists on *some* half-line *some* function $y(t)$ possessing a continuous first derivative which satisfies (2) (on that half-line).

For instance, it follows from (β*) in Section **2** that if $\int^\infty f(s)ds$ is convergent (possibly just conditionally) and, when considered as a function of its lower limit, satisfies the condition

$$\limsup_{t\to\infty} t\,\Big|\int_t^\infty f(s)\,ds\Big| < \tfrac{1}{4}, \tag{8}$$

then (1) is non-oscillatory. In particular, (1) must be non-oscillatory if

$$\int_t^\infty f(s)\,ds = o(t^{-1}) \text{ as } t\to\infty, \tag{9}$$

where the (possible just conditional) convergence of the integral is part of the assumption.

A criterion of a different type is as follows: If the $f(t)$ in (1) can be written as the sum of two functions, $f_1(t)$ and $f_2(t)$, which are such that

$$\limsup_{t\to\infty} f_1(t) < 0 \text{ and } \int^\infty f_2(t)\,dt \text{ converges}$$

(possibly just conditionally), then (1) is non-oscillatory.

In fact, $y(t) = \int_t^\infty f_2(s)\,ds$ will satisfy (2) for large t. For, since $f = f_1 + f_2$, this $y(t)$ reduces (2) to

$$[\int_t^\infty f_2(s)\,ds]^2 \leqq -f_1(t),$$

a condition which is satisfied for large t, since the integral on the left tends to 0, while $-f_1(t)$ on the right has a positive lower bound. Needless to say, this sufficient condition for non-oscillatory behavior cannot be obtained from Sturm's comparison theorem. In fact, lim sup $f_1 < 0$ and the convergence of the integral of f_2 (which, incidentally, can readily be lightened) do not preclude the possibility lim sup $f \geqq 0$ (or, for that matter, lim sup $f = \infty$) for $f = f_1 + f_2$.

5. With an arbitrary choice of the lower limit of integration, put

$$F(t) = \int^t f(s)\,ds \tag{10}$$

and suppose that

$$F(t) \to \infty \text{ as } t\to\infty \tag{11}$$

(but $f \geqq 0$ is *not* assumed). Then (1) must be oscillatory. Actually, the same is true if (11) is relaxed to

$$\int^t F(s)\,ds/t \to \infty \text{ as } t \to \infty \tag{12}$$

(cf. [5]), but the proof in the case (11) will suffice as an illustration of the use of (2) in the direction which is the opposite of that followed in Sections 2-3.

Suppose that the assertion is false. Then, although (11) is satisfied, (1) is non-oscillatory, that is, that there exists a $y(t)$ for large t. For this $y(t)$, a quadrature of (2), when combined with (10) and (11), shows that

$$y(t) + \int^t y^2(s)\,ds \to -\infty \text{ as } t \to \infty. \tag{13}$$

But (13) implies that $y(t) \to -\infty$, hence $|y(t)| \to \infty$. Consequently, the integral in (13) tends to ∞ much stronger than $|y(t)|$. Accordingly, (13) contains a contradiction.

This proves that (1) must be oscillatory if (11) holds. On the other hand, (1) need not be non-oscillatory if (11) is replaced by

$$F(t) \to -\infty \text{ as } t \to \infty. \tag{14}$$

In order to see this, suppose that $f(t)$ is periodic, and let λ denote the constant term in the Fourier series of $f(t)$. Then it is clear from (10) that (11) or (14) is satisfied according as $\lambda > 0$, $\lambda < 0$. But it is not true that $\lambda = 0$ is the cut between the oscillatory and non-oscillatory cases of (1). In fact, even Mathieu's

$$x'' + (\lambda + a^2 \cos t)x = 0, \text{ where } a \neq 0, \tag{15}$$

is known to be oscillatory for $\lambda > -\lambda^*$, where $\lambda^* = \lambda^*(a)$ is positive (cf. (23) below).

Instead of the periodicity of $f(t)$, suppose only that $f(t)$ has, for large t, a mean-value, $M(f)$. According to (10), this means that there exists a number, $M(f)$, for which

$$F(t)/t \to M(f) \text{ as } t \to \infty. \tag{16}$$

If $M(f) > 0$, then (12) is satisfied, hence (1) is oscillatory. If $M(f) < 0$, then (1) may or may not be non-oscillatory; cf. (14) and (15). But if $M(f) < 0$, and if the convergence in (16) is so rapid that the absolute value of the error does not exceed $[-M(f)]^{\frac{1}{2}}/t$ when t is large, then (1) must be non-oscillatory.

In fact, the assumption is that, if the lower limit of integration is suitably chosen, then, from a certain t onward,

$$(17) \qquad \left| \int^t f(s)\,ds - M(f)t \right| \leqq [-M(f)]^{\frac{1}{2}};$$

cf. (10) and (16). But a direct substitution shows that (2) is satisfied by

$$(18) \qquad y(t) = -\int^t f(s)\,ds + M(f)t$$

for large t, if (17) is assumed.

6. This proof also shows that if $f(t)$ is defined only on the interval $0 \leqq t \leqq 1$, and satisfies the inequality

$$(19) \quad |F(t) - F(1)t| \leqq [-F(1)]^{\frac{1}{2}}, \text{ where } F(t) = \int_0^t f(s)\,ds,\ 0 \leqq t \leqq 1,$$

then (1) is disconjugate on $I = [0, 1]$. In fact, the assumption (19), which corresponds to (17), assures that (2) is satisfied, for $0 \leqq t \leqq 1$, by the $y(t)$ which results if the lower limit of integration is chosen to be $s = 0$ in (18).

Clearly, (19) can be written in the form

$$(20) \quad \left| \int_0^t \{f(s) - \lambda\}\,ds \right| \leqq (-\lambda)^{\frac{1}{2}}, \qquad 0 \leqq t \leqq 1,\ \lambda = \int_0^1 f(t)\,dt.$$

On the other hand, an application of Schwarz's inequality shows that (20) is surely satisfied if

$$(21) \qquad \int_0^1 [f(s) - \int_0^1 f(t)\,dt]^2\,ds \leqq -\int_0^1 f(t)\,dt.$$

Finally, (21) can be interpreted by saying that the mean-value of $f(t)$ on $[0, 1]$ is non-positive and majorizes the squared standard derivation of $f(t)$. Hence, such a majorization of the standard deviation of $f(t)$ on $[0, 1]$ is sufficient to make (1) disconjugate on $[0, 1]$.

A corollary is that if $f(t)$ is periodic, of period 1, and satisfies (21), or the more general condition (20), then (1) must be non-oscillatory.

Needless to say, (20) requires a preponderance of $c_0 = -|c_0|$ in the Fourier series $f(t) \sim \sum c_n \exp(2\pi i n t)$, where $n = 0, \pm 1, \cdots$ and $c_{-n} = \overline{c_n}$. In particular, (21) is equivalent to $|c_1|^2 + |c_2|^2 + \cdots \leqq -\frac{1}{2}c_0$, by Parseval's relation.

Accordingly, if $\phi(t)$ is a continuous, periodic function of period 1 and of mean-value 0, and if λ^* denotes the cut between those λ-values for which the differential equation

(22) $x'' + \{\lambda + \phi(t)\}x = 0$, where $\phi(t) \sim \sum_{n=-\infty}^{\infty}{}' \gamma_n \exp(2\pi\, int)$, $n \neq 0$,

is oscillatory and those for which it is non-oscillatory, then

$$\lambda^* \geqq -2 \sum_{n=1}^{\infty} |\gamma_n|^2 \tag{23}$$

(and (22) is non-oscillatory also for the λ-value which is the expression on the right-hand side of (23)).

An upper limitation of λ^*, corresponding to the lower limitation (23), does not follow. All that is clear in this direction is the inequality $\lambda^* \leqq 0$, which is contained in the criterion (11).

7. In order to avoid an interruption of the proofs below, it will be convenient to isolate the following remark:

For large positive t, let $f(t)$ be defined in such a way that the improper integral

$$\int^{\infty} f(t)\,dt \text{ is convergent} \tag{24}$$

(possible just conditionally), and that (1) is non-oscillatory. Then

$$\int^{\infty} l^2(t)\,dt < \infty \tag{25}$$

must hold for the logarithmic derivative, $l = x'/x$, of every solution, $x = x(t)$, of (1). In addition,

$$l(t) \to 0 \text{ as } t \to \infty. \tag{26}$$

Finally,

$$l(t) = \int_t^{\infty} \{l^2(s) + f(s)\}\,ds. \tag{27}$$

It is understood that the lower limits of integration in (25) and (27) are supposed to be chosen large enough, larger than the last zero of $x(t)$. Since (1) is supposed to be non-oscillatory, there exists such a zero for every $x(t)$ (if there is at all a zero). Let it be denoted by t_0 (if it exists; otherwise choose t_0 arbitrarily), and let $t_0 < t < \infty$. Then $l(t) = x'(t)/x(t)$ can be formed and, as seen if (1) is divided by x, satisfies the differential equation

$$l'(t) + l^2(t) + f(t) = 0 \tag{28}$$

(Riccati). But if a quadrature is applied to (28), it follows from (24) that

$$l(T) + \int^{T} l^2(t)\,dt$$

must tend, as $T \to \infty$, to a finite limit. Hence, a repetition of the argument used after (13) shows that the negation of the truth of (25) leads to a contradiction. Consequently, (25) implies that $l(t)$ must tend, as $t \to \infty$, to a limit. Hence, (26) follows from (25). Finally, (27) follows from (26) and (28).

In the proof of (25)-(28), it was not assumed that

$$f(t) \geqq 0. \tag{29}$$

If (29) is assumed, then the hypothesis (24) can be omitted by virtue of the other hypothesis, that requiring that (1) be non-oscillatory. In other words, if (29) is assumed but (24) does not hold, then (1) must be oscillatory. In fact, (29) and the negation of (24) imply that (10) satisfies (11).

8. Without assuming (29), the following criterion will now be proved:

If (24) is assumed, and if the integral (24), when considered a function of its lower limit, satisfies the condition

$$\int^{\infty} \exp\{-2\int^{t}[\int_{u}^{\infty} f(v)\,dv]\,du\}\,dt < \infty, \tag{30}$$

then (1) must be oscillatory (the choice of the unspecified lower limits of integration in (30) is immaterial).

This criterion for the oscillatory behavior of (1) is quite different from the known criteria, since *individual* t-values do not occur in (30). In other words, $f(t)$ and its repeated indefinite integrals can be of arbitrarily "irregular growth" (even if $f(t)$ is non-negative, which is not assumed). Correponding to the fact that t occurs in (30) only beneath an integral sign, it would be inconvenient to deduce (30) from the general criterion (2). But the latter might lead to a final result, whereas it will remain undecided whether the 2 in (30) is the best absolute constant. Sure is only that the 2 cannot be more than doubled. In fact, if $f(t) = c/t^2$, then (30) is satisfied when $c > \frac{1}{2}$, whereas the case $f(t) = c/t^2$ of (1) is oscillatory if and only if $c > \frac{1}{4}$ (cf. Section **3**).

In order to prove that (1) is oscillatory under the assumptions (24)

and (30), suppose that (1) is non-oscillatory. Then, if $x(t)$ is any solution, (27) is applicable for large t. In particular, since $l = x'/x$,

$$d \log x(t)/dt > \int_t^\infty f(s)\,ds \tag{31}$$

and $x(t) \neq 0$ hold from a certain t onward. In addition, it can be assumed that $x(t) > 0$, since if $x(t)$ is a solution of (1), then $-x(t)$ is. Hence, if a quadrature is applied to (31), it follows from the assumption (30) that

$$\int^\infty [x(t)]^{-2}\,dt < \infty \tag{32}$$

holds for every solution of (1), if the lower limit of integration (34) exceeds the last zero of the solution. But this contains a contradiction, since, as shown by Hartman [3], every (non-oscillatory) differential equation (1) must have some solution violating (32).

It will be noted that this proof seems to use, via (31), just about the half of the content of (27). Actually, this is not quite correct, as seen by evaluating the integrals involved for the case $f(t) = c/t^2$ of an arbitrary $c > \frac{1}{4}$.

9. Along the lines of the preceding proof, it is easy to obtain a sufficient condition for the non-existence of an "eigensolution" (of class L^2), that is, of a solution ($\not\equiv 0$) satisfying

$$\int^\infty x^2(t)\,dt < \infty. \tag{33}$$

Suppose first that $f(t)$ satisfies (29) but (1) is non-oscillatory. Then, since $x(t)$ can be supposed to be positive for large t, it is seen from (1) that, for reasons of convexity, $x(t)$ must have a positive lower bound as $t \to \infty$, and so (33) is impossible.

Next, suppose that (29) is replaced by

$$f(t) \leqq 0. \tag{34}$$

Then (1) must be non-oscillatory (in fact, (34) implies that (2) holds for $y(t) \equiv 0$). But (1) may or may not have a solution satisfying (33), even if (24) is assumed. In fact, if $f(t) = c/t^2$, then the explicit form of the solutions, described in Section **3**, shows that (1) has a solution satisfying (33) if and only if $c < -\frac{3}{4}$.

Without asuming (34), it is easy to see that, if (1) is non-oscillatory and if $f(t)$ satisfies (24) and

$$\int^{\infty} \exp\{2 \int^{t} [\int_{u}^{\infty} f(v)dv]du\}dt = \infty, \tag{35}$$

then (1) cannot have a solution satisfying (33).

In fact, since (27) is applicable, (31) holds for every solution of (1). But the application of a quadrature on (31) shows that (33) is impossible if (35) is assumed.

For instance, a non-oscillatory differential equation (1) cannot have a solution satisfying (33) if (24) is assumed and

$$t \int_{t}^{\infty} f(s)ds \geqq -\tfrac{1}{2} \tag{36}$$

holds for large t. In fact, (36) implies that the function of t which is integrated in (35) is minorized by a constant multiple of $\exp\{2(-\frac{1}{2}\log t)\} = t^{-1}$; hence (35) is satisfied. In particular, (33) cannot hold for any solution of (1) if $f(t)$ satisfies (34), (24) and (36). It remains undecided whether the $-\frac{1}{2}$ in (36) is the best constant (whether (34) is or is not assumed). The example $f(t) = c/t^2$, mentioned after (34), shows that the $-\frac{1}{2}$ in (36) cannot be improved to any constant which is less than $-\frac{3}{4}$.

The criterion (36) is applicable only under the proviso that (1) be non-oscillatory. This proviso is sure to be satisfied if

$$t \int_{t}^{\infty} f(s)ds \leqq \tfrac{1}{4}. \tag{37}$$

In fact, (36) and (37) imply both of the inequalities (4**) of Section 2, and (4**) was seen to be sufficient in order that (1) be non-oscillatory. Accordingly, (1) cannot have a solution satisfying (33) if $f(t)$ is subject to (36) and (37) (where (24) is part of the assumption).

In particular, this will be the case if $f(t)$ is subject to (4*), since (4*) implies both (36) and (37), Still more particularly, it is sufficient to subject $f(t)$ to

$$\int_{t}^{\infty} f(s)ds = o(t^{-1}) \text{ as } t \to \infty, \tag{38}$$

since this o-condition implies (4*) for large t.

10. On a fixed t-range I, replace (1) by

$$(p(t)x')' + f(t)x = 0, \tag{39}$$

where $p(t)$ is positive and, as $f(t)$, continuous on I. Let N be defined by the property that every solution $x = x(t) \not\equiv 0$ of (39) has at most N, and some solution has actually N, zeros on I. Then $1 \leqq N \leqq \infty$. In fact, $N = 0$ is impossible, since $x(t_0) = 0$ (along with any $x'(t_0) \neq 0$) can be assigned as an initial condition at a point t_0 of I. If $N = 1$, then (39) will be called disconjugate on I.

As a generalization of the criterion (2), where $p(t) \equiv 1$, it is easy to see that (39) is disconjugate on I if and only if there exists on I *some* function $y = y(t)$ possessing a continuous first derivative and satisfying the inequality $R(y) \leqq 0$ at every point of I, where R denotes the Riccati operator associated with (39), that is,

$$R(y) = y' + y^2/p + f. \tag{40}$$

It is well-known that

(i) if $N = 1$, then (39) possesses solutions which have no zeros at all on I (Jacobi), and that

(ii) a function $y = y(t)$ having a continuous first derivative on I is a solution of the non-linear differential equation $R(y) = 0$ if and only if $y(t)/p(t)$ is the logarithmic derivative of some solution $x(t) \not\equiv 0$ of the linear differential equation (39) on I (Riccati).

Since $p(t) > 0$, and since $x(t)$ and $x'(t)$ cannot vanish simultaneously for any solution $x(t) \not\equiv 0$ of (39) on I, it follows that a solution $y(t)$ of $R(y) = 0$ cannot exist on the whole of I unless $N = 1$, and that this necessary condition, $N = 1$, is sufficient as well; cf. (i). But if $N \neq 1$, then any closed t-interval containing two zeros of a solution ($\not\equiv 0$) of (39) must contain at least one zero of any other solution (Sturm). It follows that

(iii) the non-linear differential equation $R(y) = 0$ does or does not possess some solution $y(t)$ which exists on the whole of I according as $N = 1$ or $N > 1$ (this, in turn, implies that (ii) contains (i) as a corollary).

Let (39), $R^*(.)$ and N^* denote what result if $f(t)$ in (39), (40) and the definition of N is replaced by another continuous function, $g(t)$, while $p(t)$ and I are unchanged. According to Sturm, $N \leqq N^*$ holds whenever $f(t) \leqq g(t)$ on I. If this is combined with the identity $R(.) - R^*(.) = f - g$, it follows from (i), (ii) and (iii) that $N = 1$ if and only if there

exists on I some function $y = y(t)$ possessing a continuous first derivative and satisfying $R(y) \leqq 0$ on I; q. e. d.

Remark. Let (39_*), $R_*(.)$ and N_* denote what result if $p(t)$ in (39), (40) and the definition of N is replaced by another positive, continuous function, $q(t)$, while $f(t)$ and I are left unchanged. Then $N_* \leqq N$ whenever $p(t) \leqq q(t)$ holds on I. The customary proofs of this theorem of Sturm are more elaborate than is that of his inequality $N \leqq N^*$, used above (the proof of $N \leqq N^*$ now found in textbooks depends on an *ad hoc* identity of Picone).

Actually, it is clear from the definitions of $R(.)$ and $R_*(.)$ that $R(.) \geqq R_*(.)$ whenever $0 < p \leqq q$. In view of the above criterion, this implies that (39_*) must be disconjugate on I if (39) is. This means that $N_* \leqq N$, provided that $N = 1$. But the latter proviso can readily be removed if I is subdivided in a suitable manner.

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, *Leçons sur les méthodes de Sturm*, Paris, 1917 (Collection Borel), chap. III.

[2] G. Borg, "On a Liapounoff criterion of stability," *American Journal of Mathematics*, vol. 71 (1949), pp. 67-70.

[3] P. Hartman, "On differential equations with non-oscillatory eigenfunctions," *Duke Mathematical Journal*, vol. 15 (1948), pp. 697-709.

[4] A. Kneser, "Untersuchungen über die reellen Nullstellen der Integrale linearer Differentialgleichungen," *Mathematische Annalen*, vol. 42 (1893), pp. 409-435; cf. E. Hille, "Non-oscillation theorems," *Transactions of the American Mathematical Society*, vol. 64 (1948), pp. 234-252.

[5] A. Wintner, "A criterion of oscillatory stability," *Quarterly Journal of Applied Mathematics*, vol. 7 (1949), pp. 115-117.

ON THE CLASSICAL TRANSCENDENTS OF MATHEMATICAL PHYSICS.*

By PHILIP HARTMAN and AUREL WINTNER.

Introduction. On an x-interval (a, b), let $r(x)$, $q(x)$ and the first derivative of a positive $p(x)$ be completely monotone, that is, let

$$(1) \qquad (-1)^n d^n y/dx^n \geqq 0 \text{ on } (a, b) \qquad (n = 0, 1, \cdots)$$

for $y = r, q, p'$, where it is understood that (a, b) is open and need not be bounded. It was shown in [1] that the differential equation

$$(2) \qquad p(x)y'' + r(x)y' - q(x)y = 0$$

must have a solution $y(x)$ which is completely monotone on (a, b) (here and in the sequel, the trivial solution, $y(x) \equiv 0$, is not called a solution). In the particular case $(a, b) = (0, \infty)$, it follows from the Hausdorff-Bernstein theorem that if, $r(x)$, $q(x)$ and the first derivative of a function

$$(3) \qquad p(x) > 0$$

satisfy the inequalities (1) on the half-line $x > 0$, then (2) must have a solution $y(x)$ representable in the form

$$(4) \qquad y(x) = \int_0^\infty e^{-xt}\, d\phi(t), \qquad \text{where } 0 \not\equiv d\phi(t) \geqq 0,$$

on the half-line $x > 0$. For the particular case

$$(5) \qquad x'' - q(x)y = 0$$

of (2), cf. [5].

The purpose of this paper is an application of this result to the case of certain classical linear differential equations occurring in mathematical physics. For instance, there will be considered (§ 1) the so-called toroidal functions which, in the main, are certain normalizations of the associated Legendre functions. The latter are not hypergeometric functions. Applications will, however, be given (§ 2) to the hypergeometric differential equation, and Kummer's confluent case will also be considered (§ 3).

* Received June 27, 1950.

What results in these cases for the range of an x-half-line is an integral representation of the form (4), where x is a certain elementary function of the independent variable in the standard forms of the differential equations. The results are mere existence statements, since we cannot determine the explicit form of the respective functions $\phi(t)$ in (4). But the range of the hypergeometric or confluent parameters which are treated in this manner are not those occurring in the classical integral representation of Euler and Kummer.

When the x-range is an interval contained between the two finite singular points of the hypergeometric equation, there will result (§ 2) statements on the complete monotony of certain solutions on part of that interval.

It may be mentioned that if $p > 0$, and if $y = r, q, p'$ satisfy (1) for $0 < x < \infty$, then, in general, a solution of (2), representable in the form (4), is unique up to a constant factor. For, if (2) has two linearly independent solutions of the form (4), then all solutions of (2) tend to finite limits as $x \to +\infty$. For example, if $p(x) \equiv 1$ and $q(x) > 0$ for $x \geqq X$, it follows from a criterion of Weyl [3], p. 40, that a necessary and sufficient condition in order that all solutions of (2) tend to limits is

$$\int_X^\infty \exp\Big(-\int_X^s r(\xi)\,d\xi\Big) \int_X^s q(t) \exp\Big(\int_X^t r(\eta)\,d\eta\Big)\,dt\,ds < \infty. \tag{6}$$

Thus (6) is a necessary condition in order that all solution of (2) be representable in the form (4), when $p \equiv 1$ and $q > 0$. In particular, if

$$q(x) \geqq \text{const.} > 0 \text{ for large } x, \tag{7}$$

then (6) cannot hold, and not all solutions of (2) (or their negatives) are representable in the form (4). For the integral (6) is minorized by a constant multiple of

$$\int_{X+1}^\infty \exp\Big(-\int_X^s r(\xi)\,d\xi\Big) \int_{s-1}^s \exp\Big(\int_X^t r(\eta)\,d\eta\Big)\,dt\,ds$$

$$= \int_{X+1}^\infty \int_{s-1}^s \exp\Big(-\int_t^s r(\xi)\,d\xi\Big)\,dt\,ds.$$

Since $s - 1 \leqq t \leqq s$ and since $r(x)$ is bounded as $x \to +\infty$, it follows that (6) cannot hold.

1. Toroidal functions. If in the differential equation of the associated Legendre polynomials

$$(8) \qquad (z^2-1)d^2y/dz^2 + 2zdy/dz - [\nu(\nu+1) + \mu^2/(z^2-1)]y = 0$$

the substitution

$$(9) \qquad \nu = n - \tfrac{1}{2}, \qquad \mu = m, \qquad z = \cosh x$$

is made, then there results the differential equation

$$(10) \qquad d^2y/dx^2 + (\cosh x/\sinh x)dy/dx - [n^2 - \tfrac{1}{4} + m^2/\sinh^2 x]y = 0$$

for the toroidal functions; cf., e. g., [2], p. 75.

(I) *If* $n^2 - \frac{1}{4} \geqq 0$, *then there exists a toroidal function (a solution of* (10)) *representable in the form* (4) for $0 < x < \infty$. *Furthermore, this toroidal function is unique, up to a constant factor, if and only if* $n^2 - \frac{1}{4} > 0$.

Proof of (I). A comparison of (10) and (2) shows that $p(x) \equiv 1$,

$$(11) \qquad r(x) = \cosh x/\sinh x \text{ and } q(x) = n^2 - \tfrac{1}{4} + m^2/\sinh^2 x.$$

Thus, $p'(x) \equiv 0$, $r(x)$ and $q(x)$ are completely monotone on the half-line $x > 0$ (that is, $z > 1$) if $n^2 - \frac{1}{4} \geqq 0$. In fact, since

$$(12) \qquad \cosh x/\sinh x = (1 + e^{-2x})/(1 - e^{-2x}) = (1 + e^{-2x}) \sum_{j=0}^{\infty} e^{-2jx}$$

holds for $x > 0$, it is seen that $r(x)$ is the sum of completely monotone functions and is therefore completely monotone. Similarly, the identity

$$(13) \qquad 1/\sinh x = 2e^{-x}/(1 - e^{-2x}) = 2e^{-x} \sum_{j=0}^{\infty} e^{-2jx}$$

shows that $1/\sinh x$ is completely monotone. Since the product of completely monotone functions is completely monotone, it follows that $1/\sinh^2 x$ is. Hence, $q(x)$ is completely monotone for $x > 0$ if $n^2 - \frac{1}{4} \geqq 0$. The first assertion of (I) follows from the theorem concerning (2), quoted in the Introduction.

By (11), the inequality $n^2 - \frac{1}{4} > 0$ implies that $q(x)$ satisfies (7); so that (10) has only one solution, up to a constant factor, representable in the form (4). If $n^2 - \frac{1}{4} = 0$ and $m > 0$, the general solution of (10) is

$$y(x) = c_1 \exp(m \int_1^x ds/\sinh s) + c_2 \exp(-m \int_1^x ds/\sinh s).$$

It is readily verified that this solution $y(x)$ is completely monotone for

$0 < x < \infty$ if and only if $|c_1| \leqq c_2 \exp(-2m \int_1^\infty ds/\sinh s)$. Finally, if $n^2 - \frac{1}{4} = 0$ and $m = 0$, the general solution of (10) is $y(x) = c_1 \int_1^x ds/\sinh s + c_2$, which is completely monotone if and only if $0 \leqq -c_1 \leqq c_2/\int_1^\infty ds/\sinh s$. This completes the proof of (I).

Remark. Since the change of variables $z \to -z$ leaves (9) unaltered, it follows that if $\nu(\nu+1) \geqq 0$, there exists on $-\infty < z < -1$ a solution $y = y(z)$ of (9) representable in the form (4), where $z = -\cosh x$.

2. Hypergeometric functions. The case $\mu = 0$ of (8) can be considered as a special case of the hypergeometric differential equation

$$z(1-z)d^2y/dz^2 + [c - (a+b+1)z]dy/dz - aby = 0. \tag{14}$$

In fact, if the change of independent variables $2z - 1 \to z$ is made in (14) and if $c = 1$, $a = \nu + 1$ and $b = -\nu$, then (14) becomes (8) with $\mu = 0$. Thus the results of Section 1 can be transferred from (9) to (14) if

$$c = 1, \qquad a + b = 1, \qquad ab \leqq 0. \tag{15}$$

But it turns out the corresponding statements are valid for a wider range of parameters than those specified by (15).

(II) *If the parameters satisfy*

$$ab \leqq 0 \text{ and } a + b \geqq \max(2c - 1, 0), \tag{16}$$

then (14) *possesses a solution* $y = y(z)$ *representable in the form* (4), *where* $0 < x < \infty$ *and either* $2z - 1 = \cosh x$ (*so that* $1 < z < \infty$) or $2z - 1 = -\cosh x$ (*so that* $-\infty < z < 0$). *This solution of* (14) *is unique, up to a constant factor, if and only if either* $ab < 0$ *or* $a = b = 0$.

If x is eliminated, it is seen that (II) means that (14) has, under the assumption (16), a solution of the form

$$y(z) = \int_0^\infty [2z - 1 + 2(z^2 - z)^{\frac{1}{2}}]^{-t} d\phi(t) \tag{17}$$

for $1 < z < \infty$, or of the form

$$y(z) = \int_0^\infty [1 - 2z + 2(z^2 - z)^{\frac{1}{2}}]^{-t} d\phi(t) \tag{17 bis}$$

for $-\infty < z < 0$, where $\phi(t)$ is monotone on $[0, \infty)$ and not a constant.

Proof of (II). Introduce x as a new independent variable in (14), where either $2z - 1 = \cosh x$ or $2z - 1 = -\cosh x$. The resulting differential equation in either case is

$$d^2y/dx^2 + [(a+b)\cosh x/\sinh x + (a+b+1-2c)/\sinh x]dy/dx + aby = 0. \tag{18}$$

A comparison of this differential equation with (2) shows that $p(x) \equiv 1$, $r(x) = (a+b)\cosh x/\sinh x + (a+b+1-2c)/\sinh x$ and $q(x) = -ab$. It follows therefore from (12) and (13) that

$$r(x) = (a+b)(1 + 2\sum_{j=1}^{\infty} e^{-2jx}) + 2(a+b+1-2c)\sum_{j=0}^{\infty} e^{-(2j+1)x}.$$

Thus $r(x)$ is completely monotone on $(0, \infty)$ if and only if $a+b \geqq 0$ and $a+b+1-2c \geqq 0$; that is, if and only if $a+b \geqq \max(0, 2c-1)$. Clearly, $q(x)$ is completely monotone when $ab \leqq 0$. Hence, the first assertion of (II) follows from the theorem concerning (2), quoted in the Introduction.

If $ab < 0$, then $q(x)$ satisfies (7), hence (17) has at most one solution $y(x)$ representable in the form (4). If $ab = 0$, the general solution of (17) is

$$y(x) = c_1 \int_1^x (\sinh s)^{-(a+b)} \exp[(2c-a-b-1)\int_1^s dt/\sinh t]ds + c_2.$$

If $a+b > 0$, it is easily seen that $y(x) \geqq 0$ for $0 < x < \infty$ if and only if

$$0 \leqq -c_1 \leqq c_2 / \int_1^\infty (\sinh s)^{-(a+b)} \exp[(2c-a-b-1)\int_1^s dt/\sinh t]ds;$$

in which case $y(x)$ is completely monotone. But if $a+b = 0$, then $y(x) > 0$ for $0 < x < \infty$ if and only if $c_1 = 0 \leqq c_2$. Thus the second part of (II) follows.

Integral representations distinct from (17), (17 bis) can be obtained for certain solutions of (14). In this direction, the following will be proved:

(III) *If the parameters satisfy*

$$1 + c(a+b) \geqq c + \tfrac{1}{2}(a+b)^2, \tag{19}$$

then (14) *has a solution representable in the form*

$$y(t) = (-z)^{\frac{1}{2}(1-c)}(1-z)^{\frac{1}{2}(c-a-b-1)} \int_0^\infty d\phi(t)/(1-z)^t \tag{20}$$

for $-\infty < z < 0$, *while if*

$$c + \tfrac{1}{2}(a+b)^2 \geqq c(a+b), \tag{21}$$

then (14) *has a solution representable in the form*

$$y(z) = z^{-\frac{1}{2}c}(z-1)^{\frac{1}{2}(c-a-b)} \int_0^\infty d\phi(t)/z^t \tag{22}$$

for $1 < z < \infty$, *where* $0 \neqq d\phi(t) \geqq 0$ *in both cases.*

Proof of (III). For $-\infty < z < 0$, the variation of constants and the change of independent variables determined by

$$u = (-z)^{\frac{1}{2}(c-1)}(1-z)^{\frac{1}{2}(a+b+1-c)}y, \qquad x = \log(1-z) \tag{23}$$

transform (14) into a differential equation, for u, of the form (2), where

$$p(x) = 1 - e^{-x}, \qquad r(x) = e^{-x} \tag{24}$$

and

$$4q(x) = (1-c)^2/(e^x - 1) + [1 - (a+b-c)^2]/e^x + (a-b)^2;$$

cf. [4], p. 453. Clearly, $p(x) > 0$, and $p'(x) = r(x) = e^{-x}$ is completely monotone for $0 < x < \infty$. Since $1/(e^x - 1) = e^{-x}/(1-e^{-x}) = \sum_{j=1}^{\infty} e^{-jx}$ for $0 < x < \infty$, it follows that

$$4q(x) = (a-b)^2 + [(1-c)^2 + 1 - (a+b-c)^2]e^{-x} + (1-c)^2 \sum_{j=2}^{\infty} e^{-jx}.$$

Thus $q(x)$ is completely monotone for $0 < x < \infty$ if and only if the coefficient of e^{-x} is non-negative; that is, if and only if (19) holds. If (19) is satisfied, then the differential equation for u, which resulted from (14), has a solution of the form (4), that is,

$$u(x) = \int_0^\infty e^{-xt} d\phi(t), \text{ where } 0 \neqq d\phi(t) \geqq 0 \qquad (0 < x < \infty). \tag{25}$$

In view of (23), this is equivalent to (20).

Similarly, if $1 < z < \infty$, the variation of constants and the change of independent variables determined by

$$u = z^{\frac{1}{2}c}(z-1)^{\frac{1}{2}(a+b-c)}y, \qquad x = \log z$$

transform (14) into a differential equation, for u, of the form (2), where (24) holds and

$$4q(x) = (a+b-c)^2/(e^x-1) + (2c-c^2)/e^x + (a-b)^2.$$

The remainder of the proof of the second part of (III) is similar to the proof of the first part.

The last two theorems concern certain solutions of the hypergeometric differential equation (14) on the real z-range between the singular points $-\infty, 0$ or between the singular points $1, \infty$. In some cases, it is possible to ascertain the complete monotony of solutions of (14) on the part of the interval between the singular points $0, 1$.

(IV) *If there exists a number α satisfying*

$$0 < \alpha \leqq \tfrac{1}{2}, \qquad \alpha(a+b+1) \leqq c, \tag{26}$$

and if

$$ab \geqq 0, \qquad a+b+1 \geqq 0, \tag{27}$$

then (14) *has a solution $y = y(z) \not\equiv 0$ which is completely monotone on the interval $0 < z < \alpha$.*

Proof of (IV). The coefficient, $P(z) = z(1-z) = z - z^2$, of d^2y/dx^2 in (14) is non-negative and possesses a completely monotone derivative, $P'(z) = 1 - 2z$, on the interval $(0, \frac{1}{2})$. The coefficient, $Q(z) = ab$, of $-y$ in (14) is completely monotone if $ab \geqq 0$. Finally, the coefficient, $R(z) = c - (a+b+1)x$, of dy/dx is completely monotone on $(0, \alpha)$ if the second inequality holds in both (26) and (27). In view of the theorem concerning (2), quoted in the Introduction, (IV) follows.

It should be mentioned that the result is not a consequence of standard integral representations of certain hypergeometric functions (as given, for example, by the formula near the bottom of p. 8 in [2]).

3. Confluent hypergeometric functions. Consider Kummer's differential equation,

$$zd^2y/dz^2 + (c-z)\,dy/dz - ay = 0. \tag{28}$$

A standard integral representation (cf. the last formula in [2], p. 87) shows that if

$$c > a > 0, \tag{29}$$

then (28) possesses a solution of the form (4), where $z = -x$. It turns out that (28) possesses such solutions for parameter ranges different from those specified by (29).

(V) *If the parameters satisfy*

$$a \leqq 0 \text{ and } c \geqq 0, \tag{30}$$

then (28) *possesses a solution representable in the form* (4) *for* $1 < x < \infty$, *where* $z = -x$ (*so that* $-\infty < z < 0$), *and this solution is unique, up to a constant factor, if and only if* $a < 0$; *furthermore, if*

$$0 \leqq c \leqq 2a, \tag{31}$$

then (28) *possesses a solution representable in the form*

$$y(z) = e^{\frac{1}{2}z} \int_0^\infty e^{-zt} d\phi(t), \text{ where } 0 \neq d\phi(t) \geqq 0 \qquad (0 < z < \infty), \tag{32}$$

and this solution is unique up to a constant factor.

For the case in which Kummer's normal form, (28), is replaced by that of Whittaker, cf. [6].

Proof of (V). In order to prove the first part of (V), apply the change of independent variables $z \to -x$ in (28) and write the resulting differential equation in the form

$$d^2y/dx^2 + (1 + c/x)\,dy/dx + (a/x)y = 0. \tag{33}$$

A comparison of (33) with (2) shows that $p(x) \equiv 1$, $r(x) = 1 + c/x$, $q(x) = -a/x$; so that $p > 0$, and p', r, q are completely monotone for $0 < x < \infty$, if (30) holds. Hence (20) has a non-trivial solution representable, for $0 < x < \infty$, in the form (4), where $z = -x$.

In order to ascertain the uniqueness of the solution when $a < 0$, it is sufficient to verify that (6) is violated in this case. But (6) becomes

$$\text{const.} \int_1^\infty e^{-s} s^{-c} \Big(\int_1^s e^t t^{c-1} dt \Big) ds < \infty, \qquad (X = 1),$$

and this cannot be satisfied, since the interior integral is asymptotically proportional to $e^s s^{c-1}$, as $s \to \infty$. If $a = 0$, the general solution of (33) is

$$y(x) = c_1 \int_1^x e^{-s} s^{-c} ds + c_2,$$

from which it is easily seen that $y(x)$ is completely monotone if $c_1 \leqq 0$ and if c_2 is so chosen that $y(\infty) \geqq 0$.

In order to prove the statements concerning (31) and (32), use will be made of the following variation of constants:

$$y = e^{\frac{1}{2}z} u \qquad (z = x). \tag{34}$$

This transforms (28) into

$$d^2u/dx^2 + (c/x)\,du/dx - [\tfrac{1}{4} + (2a - c)/2x]u = 0.$$

Clearly, if (31) holds, this differential equation possesses a solution representable, for $0 < x < \infty$, in the form (25). This solution is unique, up to a constant factor, since $q(x)$ satisfies (7). Finally, it follows from (34) and (25) that (20) possesses a solution of the form (32).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] P. Hartman and A. Wintner, "On the Laplace-Fourier transcendents," *American Journal of Mathematics*, vol. 71 (1949), pp. 367-372.

[2] W. Magnus and F. Oberhettinger, *Formulas and theorems for the special functions of mathematical physics*, New York, 1949.

[3] H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen*, 1909, pp. 37-63.

[4] ———, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen, II," ibid., 1910, pp. 442-467.

[5] A. Wintner, "On the Laplace-Fourier transcendents occurring in mathematical physics," *American Journal of Mathematics*, vol. 69 (1947), pp. 87-98.

[6] ———, "On the Whittaker functions $W_{km}(x)$," *Journal of the London Mathematical Society*, vol. 25 (1950), pp. 351-353.

ON THE NON-INCREASING SOLUTIONS OF $y'' = f(x, y, y')$.*

By PHILIP HARTMAN and AUREL WINTNER.

---

1. A theorem of A. Kneser [7], pp. 183-191, can be formulated as follows: Suppose that $f(x, y)$ is defined for $0 \leqq x < \infty$, $0 \leqq y < \infty$ in such a way that $f(x, y)$ is continuous; that $f(x, y) \geqq 0$; that $f(x, 0) \equiv 0$ for $0 \leqq x < \infty$; that $f(x, y)$ satisfies, with respect to y, a Lipschitz condition which is uniform in every (x, y)-rectangle; finally, that $f(x, y)$ is monotone with respect to y for every fixed $x \geqq 0$. Then, for every $y_0 > 0$, the differential equation

$$y'' = f(x, y) \tag{1}$$

has a solution $y = y(x)$ which belongs to the initial condition

$$y(0) = y_0, \tag{2}$$

exists on the half-line $0 \leqq x < \infty$, and satisfies the inequalities

$$y(x) \geqq 0 \text{ and } y'(x) \leqq 0. \tag{3}$$

It follows from a remark of Mambriani [8], p. 622, that the condition that $f(x, y)$ be monotone with respect to y can be omitted in this theorem.

The most immediate application of Kneser's theorem is the linear case,

$$y'' = q(x)y, \tag{4}$$

of (1). The above conditions are satisfied by $f(x, y) = q(x)y$, if $q(x)$ is a continuous, non-negative function for $0 \leqq x < \infty$. From this case, (4), of Kneser's theorem, it is easy to deduce a somewhat more general result; namely, that if $q(x)$ is a continuous, non-negative function for $0 \leqq x < \infty$ and $r(x)$ is an arbitrary continuous function for $0 \leqq x < \infty$, then the linear differential equation

$$y'' = r(x)y' + q(x)y \tag{5}$$

has, for every $y_0 > 0$, a solution $y = y(x)$ satisfying (2) and (3); cf. [5], p. 368.

There naturally arises the question as to whether or not Kneser's theorem

---

* Received July 2, 1950

can be generalized so as to apply to a non-linear differential equation of the form

$$y'' = f(x, y, y'). \tag{6}$$

Such generalizations are known; cf., e. g., [11], p. 623; [12], p. 105. None of these results is, however, sufficiently general to include directly the special case (5) of (6). A general theorem of this type, to be proved below, is as follows:

THEOREM I. *Let $f(x, y, z)$ be a continuous function on the octant*

$$R\colon\ 0 \leqq x < \infty, \quad 0 \leqq y < \infty, \quad -\infty < z \leqq 0 \tag{7}$$

and suppose that

$$f(x, 0, 0) \equiv 0 \quad \text{for } 0 \leqq x < \infty; \tag{8}$$

that

$$f(x, y, 0) \geqq 0 \text{ for } 0 \leqq x < \infty,\ 0 \leqq y < \infty; \tag{9}$$

finally that, for every $C > 0$, there exists a positive continuous function $\phi(z) = \phi_C(z)$ for $-\infty < z \leqq 0$, satisfying

$$|f(x, y, z)| \leqq \phi(z) \text{ for } 0 \leqq x \leqq C, \quad 0 \leqq y \leqq C, \quad -\infty < z \leqq 0 \tag{10}$$

and

$$\int_{-\infty}^{0} z\,dz/\phi(z) = -\infty. \tag{11}$$

Then there belongs to every $y_0 > 0$ a solution $y = y(x)$ of (6) which exists for $0 \leqq x < \infty$ and satisfies (2) and (3).

2. The comparatively simple nature of the proof of Theorem I will be due to a consistent application of certain results of Nagumo [9]. The proofs of Theorem I and of Theorem III below are very similar to the geometrical qualitative arguments employed in [4].

In order to prove Theorem I, a few general lemmas will be needed.

LEMMA 1. *Let $f(x, y, z)$ be a continuous function on $0 \leqq x < \infty$, $0 \leqq y < \infty$, $-\infty < z < \infty$, satisfying*

$$f(x, y, 0) > 0 \text{ for } 0 \leqq x < \infty, \quad 0 < y < \infty. \tag{12}$$

Let $y = y(x)$ be a solution of (1) on an interval $(0 \leqq)\ a \leqq x \leqq b$, belonging to initial conditions satisfying

$$y(a) > 0 \text{ and } y'(a) \geqq 0. \tag{13}$$

Then

$$y'(x) > 0 \text{ for } a < x \leqq b. \tag{14}$$

Proof of Lemma 1. The argument used in the proof will be similar to one in [13], p. 46.

It will first be shown that there cannot exist a value of x, say $x = \xi$, such that $a < \xi < b$, $y'(\xi) < 0$ (and $y(\xi) > 0$). For if such an x-value exists, let $x = \alpha (\geqq a)$ be the greatest value of x to the left of $x = \xi$ for which $y'(x) = 0$. Clearly, $y'(x) < 0$ for $\alpha < x \leqq \xi$. Hence, $y(\alpha) > y(\xi) > 0$. Consequently, (6) and (12) imply the inequality $y''(\alpha) = f(\alpha, y(\alpha), 0) > 0$. But this, together with $y'(\alpha) = 0$, contradicts $y'(x) < 0$ for $\alpha < x \leqq \xi$.

This proves a weakened form of (14),

$$y'(x) \geqq 0 \text{ for } a \leqq x \leqq b. \tag{15}$$

Since (12) implies that $y \equiv y(a)$ cannot be a solution of (6), it follows that $y'(a_1) > 0$ for some $x = a_1 (> a)$ arbitrarily near $x = a$. Arguing as above, it is seen that there cannot exist a value of $x = \xi > a_1$, where $y'(\xi) = 0$. This proves Lemma 1.

3. It is interesting to note that if (12) is weakened to (9), then Lemma 1 becomes false even if (14) is weakened to (15). This is shown by the example, $f(x, y, z) \equiv 2y \mid z \mid^{\lambda} \operatorname{sgn} z/(2 - \lambda)$, where $0 < \lambda < 1$. Correspondingly, (6) becomes

$$y'' = 2y \mid y' \mid^{\lambda} \operatorname{sgn} y'/(2 - \lambda). \tag{16}$$

If (16) is multiplied by $\mid y' \mid^{1-\lambda} (2 - \lambda)$, a quadrature gives

$$\mid y' \mid^{2-\lambda} = (y^2 - c) \operatorname{sgn} y',$$

where $c > 0$ is an integration constant. Accordingly,

$$y' = \pm \{(y^2 - c) \operatorname{sgn} y'\}^{1/(2-\lambda)},$$

where both signs, $\pm$, are admissible. If the minus sign is chosen, a quadrature gives a solution of (16), satisfying $y(0) = c^{\frac{1}{2}} > 0$, $y'(0) = 0$, but $y'(x) < 0$ for small $x > 0$.

A criterion which prevents such a possibility is as follows:

LEMMA 2. *Let* $f(x, y, z)$ *be a continuous function on* $0 \leqq x < \infty$, $0 \leqq y < \infty$, $-\infty < z < \infty$, *satisfying* (8) *and* (9). *In addition, for every* $C > 0$, *let there exist an* $M = M_C$ *such that*

$$f(x, y, z) - f(x, y, 0) \geqq Mz \text{ for } 0 \leqq x \leqq C,\ 0 \leqq y \leqq C,\ -C \leqq z \leqq C.$$

Let $y(x)$ be the same as in Lemma 1. *Then* (15) *holds. If $y'(a) > 0$, then* (14) *holds.*

Proof of Lemma 2. Suppose that (15) does not hold. Then there exists a value of x, say $x = \xi$, such that $a < \xi \leqq b$, $y(\xi) > 0$ and $y'(\xi) < 0$. Let $x = \alpha (\geqq a)$ be the greatest value of x to the left of $x = \xi$ for which $y'(x) = 0$. For $\alpha < x \leqq \xi$, put

$$q(x) = \{f(x, y(x), 0) - f(x, 0, 0)\}/y(x) \tag{17}$$

and

$$r(x) = \{f(x, y(x), y'(x)) - f(x, y(x), 0)\}/y'(x). \tag{18}$$

Then (6) shows that $y(x)$ is a solution of (5) for $\alpha < x \leqq \xi$. It follows from (8) and (9) that $q(x) \geqq 0$. The condition on f involving the constant M implies that $r(x)$ is bounded from above by M for $\alpha < x \leqq \xi$.

If a new independent variable $s = s(x)$, defined by

$$ds = \exp\left(-\int_x^\xi r(u)\,du\right)dx,$$

is introduced, then (5) becomes

$$d^2y/ds^2 = q(x)y \exp 2\int_x^\xi r(u)\,du,$$

where $x = x(s)$. Since the coefficient of y is non-negative, it follows that $y(x)$, as a function of s, is convex upwards. Hence dy/ds is a non-decreasing function of s. In terms of x, this means that

$$y'(x)\exp\int_x^\xi r(u)\,du \leqq y'(\xi) < 0$$

for $\alpha < x \leqq \xi$. Hence, by the boundedness of $r(x)$ from above and by the continuity of $y'(x)$,

$$y'(\alpha)\exp M(\xi - \alpha) < 0.$$

But this contradicts the fact that $y'(\alpha) = 0$. This proves the first assertion of Lemma 2, and the second assertion is proved in a similar manner.

4. The example (16) shows that Lemma 2 becomes false if the one-sided Lipschitz condition at $z = 0$ is weakened to a Hölder condition and the other assumptions are retained.

Lemma 1 corresponds to a uniqueness theorem ([13], p. 45) which states that if, on the (x, y, z)-space, $f(x, y, z)$ is a continuous function which is strictly increasing with y (for fixed x, z), then (6) has at most one solution joining two distinct points of the (x, y)-plane. The condition that $f(x, y, z)$ is "increasing" with y cannot be lightened to "non-decreasing" with y; cf. [13], p. 45. A simple calculation shows that if the factor sgn y' is omitted, then (16) also becomes a counter-example. What corresponds to Lemma 2 is the following uniqueness theorem:

THEOREM II. *If $f(x, y, z)$ is a continuous function on the (x, y, z)-space, is non-decreasing with respect to y (for fixed x, z) and satisfies, with respect to z, a Lipschitz condition which is uniform on every bounded portion of the (x, y, z)-space, then* (6) *has at most one solution $y = y(x)$ joining two distinct points of the (x, y)-plane.*

The proof is similar to that of Lemma 2 and will be omitted. Under the additional hypothesis that $f(x, y, z)$ has continuous partial derivatives with respect to y and z, the assertion of Theorem II was proved in [10], p. 105 and [2], p. 20.

5. For the proof of Theorem I, one more lemma will be needed.

LEMMA 3. *Let $f(x, y, z)$ satisfy the conditions of Theorem* I. *Let $x_0 > 0$ and $y_0 > 0$. Then* (6) *has on $0 \leqq x \leqq x_0$ a solution $y = y(x)$ satisfying*

$$y(0) = y_0 \text{ and } y(x_0) = 0. \tag{19}$$

Proof of Lemma 3. Let the definition of $f(x, y, z)$ be extended by placing $f(x, y, z) = f(x, y, 0)$ for $z > 0$; so that $f(x, y, z)$ is a continuous function for $0 \leqq x < \infty$, $0 \leqq y < \infty$, $-\infty < z < \infty$. The resulting differential equation (6) has a solution satisfying (19).

The last assertion follows from a result of Nagumo [9], p. 864, if slight additional conditions are imposed in $f(x, y, z)$; namely, that condition (9) be replaced by (12) and that $f(x, y, z)$ have continuous partial derivatives with respect to y and z. (The wording of Nagumo's theorem, but not the proof, requires that (8) be replaced by $f(x, 0, 0) < 0$). Actually, these additional conditions are not needed; see [14], pp. 256-258; [3], pp. 277-278 and [19], pp. 55-56. (Incidentally, the superfluity of these additional conditions can be shown by using Nagumo's result and by approximating $f(x, y, z)$ suitably, and Nagumo's Hilfssatz 1 and 2 show that these approximations

need only be carried out on some parallelepiped of the type $0 \leqq x \leqq x_0$, $0 \leqq y \leqq y_0$, $-C \leqq z \leqq C$; cf. the considerations below.)

Let $y = y(x)$ be a solution of (6) satisfying (19). In order to complete the proof of Lemma 3, it remains to show that $y(x)$ can be chosen so as to satisfy $y'(x) \leqq 0$ for $0 \leqq x \leqq x_0$. This will prove that the extension of the definition of $f(x, y, z)$ was unnecessary.

The inequality $y'(x) \leqq 0$ is clear if $f(x, y, z)$ satisfies (12). For, by Lemma 1, $y'(x) < 0$ whenever $y(x) > 0$.

If $f(x, y, z)$ does not satisfy (12), consider the differential equation

$$y'' = f(x, y, y') + \epsilon y \qquad (\epsilon > 0) \tag{6 bis}$$

instead of (6). The right-hand side of (6 bis) corresponds to the function $g(x, y, z) = f(x, y, z) + \epsilon y$. This function satisfies the analogue of (8), namely, $g(x, 0, 0) \equiv 0$; and the analogue of (12), namely, $g(x, y, 0) > 0$ for $0 \leqq x < \infty$, $0 < y < \infty$. In (10), let $C = \max(x_0, y_0)$, and let $\phi(z) = \phi_C(z)$ be defined as $\phi(0)$ for $z > 0$. Clearly,

$$\int_{-\infty} z dz/\phi(z) = -\infty \text{ and } \int^{\infty} z dz/\phi(z) = \infty,$$

while $|g(x, y, z)| \leqq \phi(z) + \epsilon y_0$ for $0 \leqq x \leqq x_0$, $0 \leqq y \leqq y_0$, $-\infty < z < \infty$. It is also clear that there exists a pair of positive constants ϵ_0, M for which

$$\int_{-M}^{-a} z dz/(\phi(z) + \epsilon y_0) > 1 + y_0 \text{ and } \int_{a}^{M} z dz/(\phi(z) + \epsilon y_0) > 1 + y_0$$

whenever $0 < \epsilon \leqq \epsilon_0$, where $a = 1 + 2y_0/x_0$. It follows from the proof of Nagumo's Satz 2, [9], pp. 864-865, that if $0 < \epsilon \leqq \epsilon_0$, then (6 bis) possesses a solution $y = y_\epsilon(x)$ satisfying (19), provided that $g(x, y, z)$ possesses continuous partial derivatives with respect to y and z. Now, $y_\epsilon'(x) \leqq 0$, by Lemma 1.

Actually, this solution satisfies

$$0 \geqq y_\epsilon'(x) \geqq -M \text{ for } 0 \leqq x \leqq x_0 \text{ and } 0 < \epsilon \leqq \epsilon_0. \tag{20}$$

For suppose, if possible, that $y_\epsilon'(x) < -M$ for some value of x, say $x = \xi$. Consider the two cases $0 \leqq \xi \leqq \frac{1}{2}x_0$ and $\frac{1}{2}x_0 < \xi \leqq x_0$. In the first case, there exists an x-value $x = x_1$ such that $y_\epsilon(x_1) = 0$, $y_\epsilon'(x_1) < 0$ and $\xi \leqq x_1 < \xi + \frac{1}{2}x_0 (\leqq x_0)$; cf. Hilfssatz 2, [9], p. 863. But then $y_\epsilon(x)$ does not exist for $x_1 \leqq x \leqq x_0$, since $f(x, y, z)$ is only defined for $y \geqq 0$. In the second case, $\frac{1}{2}x_0 < \xi \leqq x_0$, there exists an x-value $x = x_1$ such

that $y_\epsilon(x_1) = y_0$, $y_\epsilon'(x_1) < 0$ and $(0 \leqq)\ \xi - \frac{1}{2}x_0 < x_1 \leqq \xi$; cf. Hilfssatz 2, [9], p. 863. But $y_\epsilon'(x) \leqq 0$ for $0 \leqq x \leqq x_0$, so this contradicts $y_\epsilon(0) = y_0$.

Thus the proviso that $g(x, y, z)$ have continuous partial derivatives with respect to y and z can be dropped. This follows from the general theorems quoted above; for, although the majorant function $\phi(z) + \epsilon y_0$ need not satisfy

$$\int_{-\infty} z dz/(\phi(z) + \epsilon y_0) = -\infty,$$

the considerations above show that it is sufficient to consider $g(x, y, z)$ on the parallelepiped $0 \leqq x \leqq x_0$, $0 \leqq y \leqq y_0$, $-M \leqq z \leqq M$. Alternatively, as mentioned above, it is sufficient to use the Nagumo theorem and to approximate $g(x, y, z)$ suitably on this parallelepiped.

The differential equation (6 bis) and the inequalities (20) show that, by equicontinuity, there exists a sequence $\epsilon_0 > \epsilon_1 > \epsilon_2 \cdots$ having the property that $y_\epsilon(x)$, where $\epsilon = \epsilon_n \to 0$, tends to a solution $y(x)$ of (6), as $n \to \infty$. Clearly, $y = y(x)$ satisfies (19) and $y'(x) \leqq 0$ for $0 \leqq x \leqq x_0$. This proves Lemma 3.

6. *Proof of Theorem* I. Let $y_0 > 0$ be fixed and let $y = y(x, x_0)$ be a solution of (6), on $0 \leqq x \leqq x_0$, satisfying (19). If there exists, for some x_0, an x-value, say $x = \xi$, such that $0 < \xi < x_0$ and $y(\xi, x) = 0$, then it follows from $y'(x, x_0) \leqq 0$ that $y'(x, x_0) \equiv 0$ for $\xi \leqq x \leqq x_0$. Thus the solution $y(x, x_0)$ can be extended for $x_0 \leqq x < \xi$ by placing $y(x, x_0) \equiv 0$ for $x > x_0$. Hence $y = y(x, x_0)$ is a solution of (6) for $0 \leqq x < \infty$ and satisfies (2) and (3).

Consequently, it can be supposed that $y(x, x_0) > 0$ for $0 \leqq x < x_0$. Choose $x_0 = n$, where $n = 1, 2, \cdots$. It then follows from Nagumo's Hilfssatz 2 ([9], p. 863) that, for any integer $k > 0$, there exists a constant $M = M(k) > 0$ such that

$$0 \geqq y'(x, n) \geqq -M \text{ for } 0 \leqq x \leqq k \text{ and } n = k+1, k+2, \cdots.$$

By equicontinuity, there exists an increasing sequence of integers $n_1, n_2, \cdots$ such that $y(x, n_1), y(x, n_2), \cdots$ tends to a solution $y = y(x)$ of (6) for $0 \leqq x < \infty$. Clearly, $y = y(x)$ satisfies (2) and (3). This proves Theorem I.

7. The indispensability of Nagumo's condition, (10)-(11), in Theorem I is shown by the following example: Let $f(x, y, x) = 3(1 + z^2)^{1+\lambda} y^2/2\lambda$, where $0 < \lambda < 1$. Then (6) becomes

$$y'' = 3(1 + y'^2)^{1+\lambda} y^2/2\lambda. \tag{21}$$

If (21) is multiplied by $2\lambda y'/(1+y'^2)^{1+\lambda}$ a quadrature gives

$$1/(1+y'^2)^{\lambda} = c - y^3,$$

where c is an integration constant. Consider a solution $y = y(x)$ of (21) satisfying $y(0) = y_0 > 0$. For such a solution of (21), the constant c satisfies $y_0^3 < c \leqq 1 + y_0^3$. Let $y_0 > 1$, so that $c - 1 > 0$. The equation in the last formula line can be written in the form

$$y' = -\{1 - (c - y^3)^{1/\lambda}\}^{\frac{1}{2}}/(c - y^3)^{\frac{1}{2}\lambda},$$

if $y' \leqq 0$. Thus, if $y'(0) < 0$, the solution $y(x)$ continues to decrease as x increases, unless $y^3(x)$ becomes $c - 1$ for some finite value of x, say $x = \xi$.

If such a value of $x = \xi$ exists, then, by (21) and a convexity argument, $y'(x) > 0$ must hold for those $x > \xi$ for which $y(x)$ exists. But if $y_0 > 1$, and if c is an admissible integration constant, it follows from (21) and

$$\int_a^b (c - y^3)^{\frac{1}{2}\lambda}/\{1 - (c - y^3)\}^{\frac{1}{2}}\, dy < \infty, \text{ where } a^3 = c - 1 \text{ and } b^3 = c,$$

that there exists a $\xi = \xi(y_0, c)$ such that $y^3(\xi) = c - 1$. Consequently, if $y_0 > 1$, then (21) does not have a solution on $0 \leqq x < \infty$ satisfying (2) and (3).

8. There arises the need for a criterion which takes into account the complication presented by the counter-example (21). Such a criterion can be formulated as follows:

THEOREM III. *Let $f(x, y, z)$ be defined and continuous on a domain containing the octant* (7) *in its interior. Suppose that the solutions of* (6) *are uniquely determined* (*locally*) *by their initial conditions, that* (8) *holds, and that*

$$f(x, y, z) \geqq 0 \text{ on } R. \tag{22}$$

Then there exists a positive $Y(\leqq \infty)$ having the property that if $0 < y_0 < Y$, then (6) *possesses a solution on $0 \leqq x < \infty$ satisfying* (2) *and* (3)

Proof of Theorem III. Let $x_0 > 0$ be fixed and consider the solutions $y = y(x; \mu)$ of (6) determined by the initial conditions $y(x_0) = 0$ and $y'(x_0) = \mu \leqq 0$. Since (8) implies that $y(x, 0) \equiv 0$, it follows from the continuous dependence of solutions on initial conditions that, if $|\mu|$ is sufficiently small, $y(x; \mu)$ exists for $0 \leqq x \leqq x_0$. Furthermore, (22) implies that $y(x; \mu)$ is convex upwards if $\mu < 0$. Hence, $y'(x; \mu) \leqq \mu < 0$ for

$0 \leqq x \leqq x_0$. Consequently, $y(0;\mu) > 0$ if $\mu < 0$. In addition, $y(0;\mu)$ is a continuous function of μ. Hence, there exists a Y, where $0 < Y \leqq \infty$, with the property that $y_0 = y(0;\mu)$ for some μ if $0 < y_0 < Y$. In what follows, it will be supposed that y_0 has been chosen in this fashion.

Consider the solution $y = y(x, m)$ of (6) determined by the initial condition $y(0) = y_0$, $y'(0) = m$. Let Ω denote the set of negative m-values for which the solution $y = y(x, m)$ can be extended over some interval $0 \leqq x \leqq x_1$ and satisfies $y(x_1) = 0$. By the choice of y_0, the set Ω is not empty. In fact, Ω is an open set on the m-axis. For, if $y = y(x) = y(x, m)$ satisfies $y(x_1) = 0$, then $y'(x_1, m) < 0$, since $y \equiv 0$ is the only solution of (6) satisfying $y(x_0) = 0$, $y'(x_0) = 0$. It follows that, if m belongs to Ω, all numbers sufficiently close to m belong to Ω.

Let $y_0'(\leqq 0)$ be the least upper bound of Ω. It will be shown that $y = y(x) = y(x, y_0')$ exists for $0 \leqq x < \infty$ and satisfies (2) and (3).

First, if $y_0' = 0$, then $y(x)$ cannot vanish, since (22) implies $y''(x) \geqq 0$. If $y_0' < 0$, then $y(x)$ cannot vanish. For otherwise there exist values of $m > y_0'$ for which $y(x, m)$ has a zero, and this contradicts the definition of y_0'. (In particular, y_0' is a cluster point of Ω.)

Next, $y'(x)$ cannot attain a positive value for some value of x, say for $x = \xi$. For if this were possible, it would follow that there exist in Ω values of m for which $y(x, m) > 0$ for $0 \leqq x \leqq \xi$, and $y'(\xi, m) > 0$, while $y(x, m)$ vanishes for some $x_1 > \xi$. But this contradicts the inequality $y'(x, m) \leqq 0$, where $0 \leqq x \leqq x_1$. Hence, $y(x)$ satisfies (2) and (3) on its domain of existence.

The solution $y = y(x)$ can fail to exist for all $x \geqq 0$ only if there exists a positive number $x = x_2$ such that $|y(x)| + |y'(x)| \to \infty$ as $x \to x_2 - 0$; cf. [16]. But this is impossible since $0 \leqq y(x) \leqq y_0$; $y_0' \leqq y'(x) \leqq 0$, by virtue of $y''(x) \geqq 0$. This proves Theorem III.

9. The proofs of Theorem I and III seem to indicate that, if $f(x, y, z)$ is continuous and non-negative on R, and if (6) possesses a solution on $0 \leqq x < \infty$ satisfying (2) and (3), then the points $(0, y_0)$ and $(x_0, 0)$, where $x_0 > 0$, can be joined by a solution of (6). But it turns out that such is not the case.

In fact, consider the differential equation

$$y'' = -(1-y')^2 y'. \tag{23}$$

The function $f(x, y, z)$ corresponding to (23) is $f(x, y, z) = -(1-z)^2 z$, which is non-negative on R. For every y_0, the differential equation (23)

possesses the solution $y(x) \equiv y_0$ on $0 \leqq x < \infty$ satisfying (2) and (3). Considerations similar to those concerning (21) show that no solution of (23) satisfying $y(0) = y_0$ can vanish for any x if $y_0 > 1$.

10. An important question concerning the solutions of (6) which satisfy (3) for $0 \leqq x < \infty$ is whether or not the limit relation

$$\lim_{x \to \infty} y(x) = 0 \tag{24}$$

holds. In his theorem concerning $y'' = f(x, y)$, quoted in § 1, Kneser concluded (24) incorrectly; cf. [17], p. 95. For this case, Mambriani [8], p. 620, gave a sufficient condition for (24); namely, that for every $c > 0$ there should exist a positive constant $C = C_c > 0$ such that

$$f(x, y) \geqq C\sigma(x) \text{ for } c \leqq y < \infty,$$

where $\sigma(x) \geqq 0$ is a continuous function on $0 \leqq x < \infty$ satisfying

$$\int_0^\infty \sigma(x)\,dx = \infty. \tag{26}$$

This condition can be improved greatly; cf. (29) below.

The improvement will be based on the following facts: (i) if $q(x)$ is continuous for large positive x, then the linear differential equation (4) must possess a solution $y(x)$ tending to a finite, non-vanishing limit if

$$\int^\infty x\,|\,q(x)|\,dx < \infty;$$

(ii) the latter sufficient condition is necessary as well if $q(x) \geqq 0$ (in other words, if $q(x) \geqq 0$, then

$$\int^\infty xq(x)\,dx < \infty \tag{27}$$

is necessary and sufficient for the existence of a $y(x)$ for which $y(\infty)$ exists and in neither 0 nor ∞).

Assertions (i) and (ii) can be concluded from [1], § 4 and [15], § 1, respectively. A refinement of (i) and a simple proof of (ii) are given in [18].

THEOREM IV. *Let $f(x, y, z)$ be a continuous function on R, with the property that, for every pair of constants $C > c > 0$, there exists a non-negative, continuous function $\sigma(x) = \sigma(x; c, C)$ satisfying*

$$f(x, y, z) \geqq \sigma(x) \text{ for } 0 \leqq x < \infty,\ c \leqq y \leqq C,\ -C \leqq z \leqq 0 \tag{28}$$

and

$$\int^{\infty} x\sigma(x)\,dx = \infty. \tag{29}$$

Then (24) *holds if* $y = y(x)$ *is any solution of* (6) *on* $0 \leqq x < \infty$ *which satisfies* (3).

Proof of Theorem IV. Since the functions $\sigma(x) = \sigma(x; c, C)$ are non-negative, it follows from (28), (22) and (6) that $y''(x) \geqq 0$. Hence, if $C \geqq -y'(0)$, it is clear that $-C \leqq y'(x) \leqq 0$ for $0 \leqq x < \infty$.

Suppose that the assertion of Theorem IV is false. Then $y(\infty) = c > 0$ holds for some $y(x)$. Hence, if $C \geqq y(0)$, then $c \leqq y(x) \leqq C$ for $0 \leqq x < \infty$. Choose $c = y(\infty) > 0$ and $C = \max(y(0), -y'(0))$.

By (6), the function $y = y(x)$ satisfies the linear differential equation (4), where $q(x) = f(x, y(x), y'(x))/y(x)$. It is seen from (28) that $q(x) \geqq \sigma(x)/y(x) \geqq \sigma(x)/c$. Hence (29) implies that (27) cannot hold. Consequently, $y(\infty) = c > 0$ is impossible. This contradiction proves Theorem IV.

11. If (4) is replaced by (5), then (27) becomes

$$\int_0^{\infty} \{\int_0^{u} \exp(-\int_u^{u} r(s)\,ds)\,du\} q(x)\,dx < \infty \tag{30}$$

(which reduces to (27) if $r(x) \equiv 0$). By taking advantage of this fact, one can prove a theorem similar to Theorem IV. In contrast to the linear case, the resulting theorem is not more general than Theorem IV.

THEOREM IV*. *Let* $f(x, y, z)$ *be a continuous function on* R, *with the properties that* (8) *holds and that, for every pair of constants* $C > c > 0$, *there exists a pair of continuous functions,* $\sigma(x) = \sigma(x; c, C)$ *and* $\tau(x) = \tau(x; c, C)$, *satisfying*

$$f(x, y, 0) \geqq \sigma(x) \geqq 0 \text{ for } 0 \leqq x < \infty,\ c \leqq y \leqq C; \tag{31}$$

and

$$f(x, y, z) - f(x, y, 0) \geqq \tau(x)z \text{ for } 0 \leqq x < \infty,\ c \leqq y \leqq C,\ -\infty < z < 0; \tag{32}$$

finally,

$$\int_0^{\infty} \{\int_0^{u} \exp(-\int_u^{u} \tau(s)\,ds)\,du\} \sigma(x)\,dx = \infty. \tag{33}$$

Then (24) *holds if* $y = y(x)$ *is any solution of* (6) *on* $0 \leqq x < \infty$ *which satisfies* (3).

It is not assumed that $\tau(x) \geqq 0$ in (32). Incidentally, if $f(x, y, z)$ satisfies either (22) or Nagumo's condition (10)-(11), then $y'(x)$ is bounded (cf. Hilfssatz 1, [9], p. 862). In this case, the range $-\infty < z < 0$ in (32) can be replaced by $-C \leqq z < 0$.

Proof of Theorem IV*. Suppose that the assertion of Theorem IV* is false. Then $y(\infty) = c > 0$ for some $y(x)$. This solution $y = y(x)$ of (6) cannot be constant for large values of x. For otherwise $f(x, c, 0) \equiv 0$ for large x, and this contradicts (31) and (33). Since Lemma 2 is applicable, it follows that $y'(x) < 0$ for $0 < x < \infty$.

Since $y = y(x)$ satisfies (6), it satisfies (5) if $q(x)$, $r(x)$ are defined by (17), (18). Then $q(x) \geqq \sigma(x)/y(x) \geqq \sigma(x)/c$, by (8) and (31). But $r(x) \leqq \tau(x)$ by (32). Hence the factor $\{\cdots\}$ in (30) is not less than that in (33). Thus (33) shows that (30) cannot hold. Since this implies that $y(\infty) = 0$, Theorem IV* follows.

12. By combining Theorem III with a procedure applied in [17] and [5] to the case of a linear differential equation, the following theorem will now be proved:

THEOREM V. *Let* $g(x, y, z)$ *be defined and of class* C^∞ *on* R, *and suppose that*

$$g(x, 0, 0) \equiv 0 \text{ for } 0 \leqq x < \infty \tag{34}$$

and that the partial derivatives of g *satisfy the inequalities*

$$(-1)^{i+k}\partial^{i+j+k}g/\partial x^i \partial y^j \partial z^k \geqq 0 \text{ for } i, j, k = 0, 1, \cdots \text{ on } R. \tag{35}$$

Let $p(x) > 0$ *be a continuous function on* $0 \leqq x < \infty$, *and suppose that* $p(x)$ *possesses a completely monotone first derivative* $p'(x)$ *on* $0 < x < \infty$, *that is,*

$$(-1)^{n-1} d^n p/dx^n \geqq 0 \text{ for } n = 1, 2, \cdots \tag{36}$$

$(0 < x < \infty)$. *Then there exists a positive* $Y (\leqq \infty)$ *such that, if* $0 < y_0 < Y$, *the differential equation*

$$p(x)y'' = g(x, y, y') \tag{37}$$

possesses a solution $y = y(x)$ satisfying (2) *and representable in the form*

$$y(x) = \int_0^\infty e^{-xt} d\psi(t) \text{ for } 0 < x \leqq \infty \tag{38}$$

where $\psi(t)$ is non-decreasing for $0 \leqq t < \infty$.

Proof of Theorem V. First, $f(x, y, z) = g(x, y, z)/p(x)$ satisfies the conditions of Theorem III. In fact, (8) follows from (34), and (22) from the case $i = 0$, $j = 0$, $k = 0$ of (35). Clearly, the definition of $f(x, y, z)$ can be extended, to a domain which contains R in its interior, in such a way that f_y and f_z continue to exist and to be continuous.

Thus Theorem III is applicable. Hence there exists a positive $Y(\leqq \infty)$ with the property that (37) possesses, if $0 < y_0 < Y$, a solution satisfying (2) and (3). It is clear that $y(x)$ is of class C^∞ for $0 < x < \infty$.

In order to prove that $y(x)$ has a representation of the form (38), it is sufficient to show that $y(x)$ is completely monotone for $0 < x < \infty$:

$$(-1)^n d^n y/dx^n \geqq 0 \text{ for } n = 1, \cdots \text{ if } 0 < x < \infty \tag{39}$$

(Hausdorff-Bernstein). The inequalities (3) are the cases $n = 0, 1$ of (39). The cases $n \geqq 2$ follow by induction from (3), (35) and (36), by successive differentiations of (37); cf. the arguments in [6], pp. 350-351. This proves Theorem V.

Remark. The function $f(x, y, z)$ occurring in the preceding proof satisfies the conditions of Theorem I (which assures that $Y = \infty$) if and only if

$$\partial^3 g(0, y, z)/\partial z^3 \equiv 0 \text{ for } 0 < y < \infty, \ -\infty < z \leqq 0. \tag{40}$$

In order to see this, note that (34) and (35) imply that (8) and (9) hold, and that

$$0 \leqq f(x, y, z) \leqq f(0, C, z) \text{ for } 0 \leqq x \leqq C, \ 0 \leqq y \leqq C, \ -\infty < z \leqq 0.$$

Furthermore, $f(0, C, z) = g(0, C, z)/p(0)$ is a non-increasing function of z for $-\infty < z \leqq 0$. Hence Nagumo's condition (10)-(11) can be satisfied if and only if

$$\int_{-\infty} z dz/g(0, C, z) = -\infty \text{ for every } C > 0. \tag{41}$$

The inequalities (35) imply that

$$(-1)^k \partial^k g(0, C, z)/\partial z^k \geqq 0 \text{ for } k = 0, 1, \cdots \text{ if } -\infty < z \leqq 0. \tag{42}$$

Hence, unless

$$\partial^3 g(0, C, z)/\partial z^3 \equiv 0 \text{ for } -\infty < z \leqq 0, \tag{43}$$

it follows that

$$g(0, C, z) \geqq \text{const.} \mid z \mid^3, \text{ where const.} > 0. \tag{44}$$

This implies that (41) cannot hold.

In order to see that the negation of (43) implies (44), suppose that $\partial^3 g(0, C, z_0)/\partial z^3 \leqq \text{Const.} < 0$ for some z, say $z = z_0 \leqq 0$. Then

$$\partial^3 g(0, C, z)/\partial z^3 \leqq \text{Const.} < 0 \text{ for } -\infty < z \leqq z_0,$$

by the case $k = 4$ of (42). This proves (44). Hence, (41) cannot hold. On the other hand, it is clear that (43) is sufficient for (41). This proves the Remark concerning (40).

THE JOHNS HOPKINS UNIVERSITY.

REFERENCES.

[1] M. Bôcher, "On regular singular points of linear differential equations of the second order whose coefficients are not necessarily analytic," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 40-52.

[2] R. Caccioppoli, "Problemi non-lineari in analisi funzionale," *Rendiconti del Seminario della R. Università di Roma*, ser. 3, vol. 1 (1933), pp. 13-22.

[3] S. Cinquini, "Sopra i problemi di valori al contorno per equazioni differenziali del secondo ordine," *Annalo della R. Scuola Normale Superiore di Pisa*, ser. 2, vol. 8 (1939), pp. 271-283.

[4] P. Hartman and A. Wintner, "On the asymptotic behavior of the solutions of a non-linear differential equation," *American Journal of Mathematics*, vol. 68 (1946), pp. 301-308.

[5] ——— and A. Wintner, "On the Laplace-Fourier transcendents," *ibid.*, vol. 71 (1949), pp. 367-372.

[6] ——— and A. Wintner, "On non-linear differential equations of first order," *ibid.*, vol. 72 (1950), pp. 347-358.

[7] A. Kneser, "Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I," *Journal für die reine und angewandte Mathematik*, vol. 116 (1896), pp. 178-212.

[8] A. Mambriani, "Su un teorema relativo alle equazioni differenziali ordinarie del 2º ordine," *Atti della Reale Accademia Nazionale dei Lincei*, ser. 6, vol. 9 (1929), pp. 620-622.

[9] M. Nagumo, "Ueber die Differentialgleichung $y'' = f(x, y, y')$," *Proceedings of the Physico-Mathematical Society of Japan*, ser. 3, vol. 19 (1937), pp. 861-866.

[10] A. Rosenblatt, "Sur les théorèmes de M. Picard dans la théorie des problèmes aux limites des équations différentielles ordinaires non linéaires," *Bulletin des Sciences Mathématiques*, ser. 2, vol. 57 (1933), pp. 100-106.

[11] G. Scorza-Dragoni, "Su un'equazione differenziale particolare," *Atti della Reale Accademia Nazionale dei Lincei*, ser. 6, vol. 9 (1929), pp. 623-625.

[12] ———, "Il problema dei valori ai limiti studiato in grande per gli integrali di una equazione differenziale del secondo ordine," *Giornale di Mathematiche*, ser. 3, vol. 69 (1931), pp. 77-112.

[13] ———, "A propositio di alcuni teormi relativi ad un problema ai limiti per una equazione differenziale del secondo ordine," *Atti della Reale Accademia Nazionale dei Lincei*, ser. 6, vol. 22 (1935), pp. 44-48.

[14] ———, "Elementi uniti di trasformazioni funzionali e problemi di valori ai limiti," *Rendiconti del Seminario Matematico della R. Università di Roma*, ser. 4, vol. 2 (1938), pp. 255-275.

[15] H. Weyl, "Ueber gewöhnliche lineare Differentialgleichungen mit singulären Stellen und ihre Eigenfunktionen," *Göttinger Nachrichten*, 1909, pp. 37-63.

[16] A. Wintner, "The infinities in the non-local existence problem of ordinary differential equations," *American Journal of Mathematics*, vol. 68 (1946), pp. 173-178.

[17] ———, "On the Laplace-Fourier transcendents occurring in mathematical physics," *ibid.*, vol. 69 (1947), pp. 87-98.

[18] ———, "On almost free linear motions," *ibid.*, vol. 71 (1949), pp. 595-602.

[19] G. Zwirner, "Un criterio di esistenza per un problema di valori al contorno per equazioni differenziali del secondo ordine," *Rendiconti del Seminario Matematico della R. Università di Padova*, vol. 10 (1939), pp. 55-64.

THE TOPOLOGY OF REGULAR CURVE FAMILIES WITH MULTIPLE SADDLE POINTS.* [1]

By WILLIAM M. BOOTHBY.

Introduction. It is known that the level curves of any function $f(x, y)$ which is harmonic in a simply connected domain form a curve family which is regular (locally homeomorphic to parallel lines) in the neighborhood of every point of the domain with the exception at most of a set of isolated points at each of which the curve family has a singularity of the multiple saddle point type (Figure 1). The central task of this and a later paper [2] is to

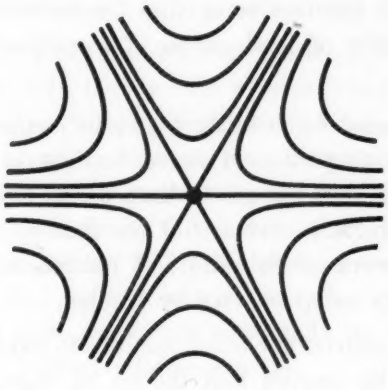

FIGURE 1
MULTIPLE SADDLE POINT

prove that these local properties are sufficient to characterize topologically the level curves of such harmonic functions, and to develop certain general topological properties of these families of curves which should prove useful in the study of harmonic functions and analytic functions. These investi-

* Received January 30, 1950.

[1] The material in this paper and the one to follow was taken from the author's Ph. D. thesis at the University of Michigan. The author wishes to express his gratitude to Professor Wilfred Kaplan for his guidance in this research and his advice in the preparation of this paper.

[2] The Topology of the Level Curves of Harmonic Functions with Critical Points, to appear in this JOURNAL.

gations generalize some of the results of several papers of W. Kaplan in which curve families which were regular (without singularities) throughout a simply connected domain were considered. In particular it was proved by Kaplan that (1) every such curve family is homeomorphic to the level curves of a function harmonic in a simply connected domain [VIII]; (2) every such family is homeomorphic to the solution family of a system of differential equations $dx/dt = f(x, y)$, $dy/dt = g(x, y)$ [VII]; and (3) that every such family can be decomposed into the sum of a denumerable collection of non-overlapping subfamilies each homeomorphic to the parallel lines of a half-plane. These results are all extended in this and the following paper to the more general type of curve family with isolated singularities of the multiple saddle point type. More precisely, the generalization of (1), (2), and (3) is achieved in the following paper; in this paper the necessary preliminary theorems are proved about the general topological structure of such families, the principal theorem being that the curve family can always be considered as the family of level curves of a continuous function without relative extrema.

Section **1** is devoted to definitions and to enumeration without proof of certain important properties of curve families in general. The most important of these deal with the sets of limit points $L(C+)$, $L(C-)$, $L(C)$, of the individual (directed) curves C of the family. These theorems stem essentially from the work of Poincaré and Bendixson. They depend, however, on the regularity only, as noted by Kaplan.

In Section **2** we restrict our curve families to have isolated singularities only and note that this implies that the set of singularities must then be closed and denumerable. The important notion of index (relative to the curve family) of a closed curve and of a singularity is defined and the theorems we will need stated, in particular the fact that the index of a closed curve is equal to the sum of the indices of the singularities in its interior. This notion of index was defined for curve families by Hamburger [see IX] and it coincides with the Poincaré index when the curve family is the family of integral curves of a vector field. It was studied in detail by Y. S. Chin [II] from which source the theorem mentioned was taken.

Finally in Section **3** we restrict our curve families to be regular in a simply connected domain except for multiple saddle points, which, for reasons explained below, will henceforth be called branch points. Such a family will be called a branched regular curve family and is the principal object of interest in what follows. The singularities will by their very nature be isolated, hence

form a closed denumerable set. The index forms a useful tool for ruling out many undesirable configurations of curves, such as closed curves, "polygons" of curves (with singular points as vertices), and so on. Moreover it enables us to investigate some of the relations between the curve family and cross-sectional arcs (i. e. arcs which locally intersect each curve only once). Finally it enables us to show that each curve either extends to infinity in a given direction on it or ends at a branch point.

From this it follows easily that the curves are of two types, one the regular curves which extend to infinity in each direction, the other branched curves which end in a branch point at one or both ends. These latter then naturally form configurations which are trees (without ends) in the topological sense, i. e. they are locally finite, one dimensional simplicial complexes, with curves of the family as 1-cells, branch points as 0-cells. It is shown by use of a theorem of Adkisson and MacLane [I] that any such tree may be straightened out to a rectilinear one by a homeomorphism of the entire plane.

Finally in Section **4** it is shown that the trees are distributed in such a fashion on the plane that, if it is done with sufficient care, we may remove certain curves of each tree with their endpoints, i. e. make cuts in the plane extending to infinity along the tree, in such a fashion that the plane with these curves removed is simply connected and the remaining curves of the family form a regular curve family filling this simply connected subset of the plane. It follows from Kaplan [IV] that these curves are the level curves of a continuous function without relative extrema, and by extending this function to the entire plane, we obtain our family too as the level curves of a continuous function without relative extrema.

1. General Properties of Regular Curve Families in the Plane.

1.1. Curve families filling a region. An *open curve* will mean a homeomorphic image of an open interval, a *closed curve* a homeomorphic image of a circle, a *half-open curve* a homeomorphic image of a half-open interval, and an *arc* of a closed interval. A *curve* will mean any one of these four. A family F of curves will be said to *fill* a subset R of the Euclidean plane π if every curve of F is in R and every point of R lies on one and only one curve of F. If U is a subset of π such that each curve C of a family F filling R intersects U in a set $U \cap C$ each of whose components is a curve, then we denote by $F[U]$ the curve family filling $U \cap R$ whose curves are the components of $C \cap U$ for all C in F. If the curve family F fills R and the curve family G fills S, then F and G will be called *homeomorphic* if there is a

homeomorphism of R onto S such that the image of each curve in F is a curve in G. If p is a point of R, R filled by a curve family F, then C_p will denote the curve of F through p.

1.2. Regularity. Henceforth, F will denote a curve family filling a subset R of π, the *oriented* Euclidean plane. We shall use R_0 to denote the rectangle $|x| \leqq 1$, $|y| \leqq 1$, of the xy-plane, and F_0 the family of lines $y =$ constant filling R_0. F will be said to be *regular* at a point p of R if there is a set $U(p)$ in R to which p is interior (relative to R) and such that $F[U(p)]$ is homeomorphic to F_0. F is then *regular in* R if F is regular at every point of R. A *cross-section* of F (*through* the point r) is an arc pq (to which r is interior), such that pq lies in a subset R' of R which is open relative to R, and such that each curve of $F[R']$ meets pq at most once. An *r-neighborhood* of a point p of R will mean a subset $U(p)$, $\overline{U(p)} \subset R$, which contains p and is open relative to R, and is moreover such that $F[\overline{U(p)}]$ is homeomorphic to the family F_0, filling the rectangle R_0 (above) of the xy-plane, in such a way that the inverse images of the lines $|x| = 1$ are cross-sections. The following properties of regular curve families will be useful later.

THEOREM 1.2-1. *If a family F fills an open region R and is regular in R, then each curve of F is either open or closed in π.* [IV, 1]

THEOREM 1.2-2. *If a family F fills any region R and is regular in R, then every point p of R has an arbitrarily small r-neighborhood $U(p)$, and there is a cross-section pq with p as endpoint. If p is in the interior of R, then there is a cross-section through p. Moreover if st is any arc lying on a curve C of F, then there is, within any ϵ-neighborhood $U_\epsilon(st)$, an r-neighborhood containing st.* [IV, 8 and VI, 1]

THEOREM 1.2-3. *Let F be a curve family filling R where $R = R_1 \cup R_2$ and $F[R_1]$ and $F[R_2]$ are both defined. If p is an interior point of R, and $F[R_1]$ and $F[R_2]$ are both regular at p, then F is regular at p.* [VI, 2]

1.3. The sets $L(C+)$ and $L(C-)$. From this point on we assume every curve of F directed, although we may later find it at times convenient to reverse directions on a given curve. If C is any open curve in F, then by a *positive* (*negative*) *limit point* of C will be meant any point q which is the limit of a sequence $p_n = f(t_n)$, where C is the image under f of $0 < t < 1$ with increasing t corresponding to the positive direction on C and $t_n \to 1$ ($t_n \to 0$). The set of all positive (negative) limit points of the directed

curve C will be denoted by $L(C+)$ (by $L(C-)$). $L(C)$ is defined by $L(C) = L(C+) \cup L(C-)$ and $C \cup L(C)$ is the point set closure of C. Clearly, $L(C) \cap C$ is empty since C is homeomorphic to $0 < t < 1$. The theorems below involving $L(C+)$ hold equally well for $L(C-)$.

THEOREM 1.3-1. *If C is an element of a regular curve family F, and $L(C+)$ contains a closed curve D of F, then $L(C+) = D$.* [IV]

THEOREM 1.3-2. *If F is a regular curve family filling R and p is in R and, moreover, in $L(C+)$ for some curve C of F, then every point of the curve D_p of F through p is in $L(C+)$.* [IV, 7]

2. Isolated Singularities and Index.

2.1. Isolated singularities. Let F be a curve family regular in R and b be an isolated boundary point of R. Then, if for some neighborhood $U(b)$ in π, F is regular in $U(b) - b$ but not in $U(b)$, we shall call b an *isolated singularity* of F. In [VI] W. Kaplan has completely classified the structure of a regular curve family in the neighborhood of an isolated singularity. Below we shall actually be interested in a very restricted type of isolated singularity; however, in this section no restrictions will be made.

THEOREM 2.1-1. *Let F be a regular curve family filling the region R, where R consists of the entire plane minus a set of points B, each of which is an isolated singularity of F. Then B is a closed and denumerable point set.*

Proof. From the fact that each point of B has a neighborhood containing no other point of B it follows at once that B is denumerable and that no point of B is a limit point of B. Now consider a point p at which F is regular. We must show that p is not in $\overline{B}$, whence $\overline{B} = B$. By the definition of regularity there is a set $U \subset R$ such that p is interior to U relative to R and there is a homeomorphism $f: U \to R_0 = \{(x, y) \mid |x| \leqq 1; |y| \leqq 1\}$. It is clear that U is a Jordan domain bounded by the inverse image of the closed curve which forms the boundary of R_0. Moreover we can, as is well-known, extend f to map all of π on the xy-plane. Hence, for simplicity, we shall deal with the homeomorphic image of π, R, F, etc. on the xy-plane. If $p' = f(p)$ is interior to $R_0 = f(U)$, then p' is not in $\overline{B'} = \overline{f(B)} = f(\overline{B})$ since the complete description of U, the r-neighborhood of p, makes it clear that the interior of U (in π) is in R. Now assume p' lies on the boundary of R_0. Since p' is by definition of U, an interior point of R_0 relative to R', there is a small circle which with its interior intersects R' in a set contained in R_0. Then clearly, if this circle is of radius less than one, at least two

quadrants of it lie outside R_0 hence outside R', i. e. in B'. It follows that B' cannot be a set of isolated points. Therefore our assumption that there was such a point on the boundary of R_0 is false, and p' is interior to R_0 and not in $\bar{B}'$. Hence $\bar{B}' = B'$ and $B = B$, which completes the proof.

2. 2. Index. Following the definition given in Kerékjártó [IX, p. 251 ff.], we define the *index* of an isolated singularity b of a family F as follows: Let K be any simple closed curve containing the isolated singularity but no other singularities on it or in its interior and let $U_1, \cdots, U_n$ be a covering of K by r-neighborhoods. Then, if the neighborhoods have been chosen sufficiently small, it is clear that we may replace K by a simple closed curve K' in $\bigcup_{i=1}^{n} U_i$ and such that (i) K' is a "polygon" composed of sides which are alternately (1) arcs of curves of F and (2) cross-sections of F, and (ii) K, K' each contain in their interiors only the singularity b. Every vertex of the polygon K' is the intersection of a cross-section and a curve of F; we call it an *internal* vertex, if the curve of F which forms the side at that vertex enters the interior of K' at the vertex, and in the other case we call it an *external* vertex. If we denote the number of internal vertices by e and external vertices by a, then let $\rho(K') = 1 - (a - e)/4$, it is shown by Kerékjártó that $\rho(K')$ so defined is independent of the particular K' chosen, depending only on the singularity b. We may then denote this number by $\rho(b)$ and define it to be the index of b.

More generally, given any closed curve K in the region R filled by the regular curve family F, we may in the same fashion define a curve K' which is close to K, which contains in its interior exactly those singularities contained interior to K, and which consists itself of alternate arcs and cross-sections. Exactly as above we may define $\rho(K') = 1 - (a - e)/4$ and show that this quantity depends only on K, not on the choice of K'. Hence we denote it $\rho(K)$ and call it the index of the curve K. Since we have restricted ourselves to families with isolated singularities, there will of course be only a finite number of singularities interior to any closed curve. The index of a closed curve relative to a given curve family has been studied in detail by Y. S. Chin [II] and he proves the following:

THEOREM 2. 2-1. *Let K be a simple closed curve in R, then $\rho(K) = \sum_{i=1}^{n} \rho(b_i)$ where b_i, $i = 1, \cdots, n$ are the singularities of F interior to K.*

This theorem will not be proved since a detailed proof may be found in [II]. We also state the following useful theorem from Kerékjártó.

THEOREM 2.2-2. *Let b be an isolated singularity of the regular curve family F and let b be the positive endpoint of exactly p curves of F and the negative endpoint of exactly q curves of F. Then* $\rho(b) = 1 - (p+q)/2$.

3. Regular Curve Families Whose Singularities Are Branch Points.

3.1. Branch points. Let b be a boundary point of $R \subset \pi$ and F a curve family regular in R. If there exists a neighborhood $U(b)$ such that $F[U(b) - b]$ is homeomorphic to the level curves of $\Re(z^n)$, the real part of $f(z) = z^n$, $n > 1$, under a homeomorphism g carrying $U(b)$ onto $|z| < 1$ with $g(b) = 0$, then we say that b is a *branch point* of F with multiplicity n. This configuration is probably better known as a multiple saddle point, but the term branch point will be adopted below for two reasons: first, these points appear as branch points on trees and, second, they will have branch points of the inverse function as images if we are concerned with the level curves of the real part of an analytic function. The term saddle point was used in the title because it seems to be more suggestive of the type of curve family considered. (See Figure 1.)

The neighborhood $U(b)$ above will be called an *admissible* neighborhood; more precisely $U(b)$ plus the homeomorphism g, which shall always be taken as an orientation preserving homeomorphism, will constitute an admissible neighborhood. In the case of a branch point b of multiplicity n, there are exactly $2n$ curves of $F[U(b)]$ which end at b, i. e. which may be directed so that $L(C+) = b$. It follows that the multiplicity is independent of the choice of $U(b)$. It is clear that b is an isolated singularity of F.

From now on we shall consider only curve families F which are regular in a region R which in general will contain all of π except for a set B of branch points of F. At times it may be expedient to let $R \cup B$ be not all of π but a simply connected domain in π. Of course, topologically the two are equivalent and for convenience we shall consider almost always the first case. Such families will be called *branched regular curve families* filling[3] π (or an open, simply-connected subset of π). The following theorem is well-known:

THEOREM 3.1-1. *The family F of level curves of a function $f(x, y)$ harmonic in a simply connected*[4] *domain D is a branched regular curve family filling D. The critical points of $f(x, y)$ are the branch points of F.*

---

[3] Strictly speaking, according to the definition in 1.1, the curve family fills not π but π *minus the branch points.* For convenience the latter phrase will often be dropped below.

[4] Simple connectedness is not needed for this theorem, but is needed for the converse.

It is this theorem which motivates in large part this study of such curve families. A converse to this theorem will be proved in a subsequent paper.

3. 2. General properties of branched regular curve families. Restricting ourselves now to branched regular curve families F filling π, i. e. more precisely $R = \pi - B$, B the set of branch points, we make a number of preliminary definitions. The curves of F are assumed directed, so that there shall be no ambiguity in the use of the symbols $L(C+)$ and $L(C-)$. We note that since R is open by Theorem 2. 1-1, every curve of F is by Theorem 1. 2-1 either open or closed. If $L(C) = 0$ we call C a regular curve, and if $L(C+) = b \in B$, i. e., a branch point, we shall say that C is a *branched curve,* branched at the positive end at b; we also call b the *positive endpoint* of C in this case. Similarly if $L(C-) = b' \in B$. It will subsequently be shown that the only possibility is that $L(C+) = 0$ or $= b$, a single branch point, and similarly $L(C-)$, so we shall not give any name to the as yet possible type of curve which might have more than one positive (or negative) limit point).

If $b \in B$, the curves of F which have b as endpoint, together with their endpoints, form a set called the *star of* b, $St(b)$; and the same curves without their endpoints, except b, form the *open star of* b. If b is of multiplicity n, then there are at most $2n$ curves in $St(b)$; it will be shown later in this section that there are exactly $2n$, i. e., the two endpoints of a curve cannot coincide. It is useful to note that since $St(b)$ contains a finite number of curves, plus the fact that B is denumerable, there are at most a denumerable number of branched curves in F.

If $C_1, \cdots, C_n$ are $n \geqq 2$ distinct branched curves of F with their endpoints, and may be so directed that $L(C_i+) = b_i = L(C_{i+1}-)$, $b_i \in B$ and b_i distinct for $i = 1, \cdots, n-1$, and if in addition $L(C_1-) \neq b_i \neq L(C_n+)$, $i = 1, \cdots, n-1$, then we call $C_1, \cdots, C_n$ either a *simple polygon* of branched curves or a *chain* of branched curves according to whether or not there is a $b_0 \in B$ such that $L(C_1-) = b_0 = L(C_n+)$. In brief, the curves $C_1, \cdots, C_n$, together with their endpoints will form a chain if the collection is homeomorphic to an arc, and a simple polygon if homeomorphic to a circle. For convenience we shall call a single curve a polygon if it has coincident endpoints. The theorem which follows shows that the polygon configuration cannot occur in F.

THEOREM 3. 2-1. *A branched regular curve family F filling π can contain neither a closed curve nor a simple polygon of branched curves.*

Proof.[5] Suppose F contains a closed curve C. Kaplan has shown [IV] that if a regular curve family contains a closed curve, the interior of the curve must contain a boundary point, i. e., in this case an element of B. Thus it follows from Theorem 2. 2-2 that the sum of the indices of the branch points interior to the closed curve is a negative integer, since by that theorem the index of a branch point of multiplicity n is $1 - n$. However, the index of the closed curve itself is one, which is a contradiction to Theorem 2. 2-1. Thus F contains no closed curve.

Next we suppose that F contains a polygon K of branched curves. We construct a curve family filling the circular region $R = \{(x, y) \mid x^2 + y^2 \leqq 4\}$ of the xy-plane as follows: fill the annular region $R_1 = \{(x, y) \mid 1 < x^2 + y^2 \leqq 4\}$ with circles $x^2 + y^2 = c^2$, $1 < c \leqq 2$, then map K with its interior onto R_2, the unit circular disc, K mapping onto the circumference. Then $R = R_1 \cup R_2$, except for the images of the singular points interior to K plus certain of the images of those on K, is filled by a curve family F'. $F'[R_1 - S]$ and $F'[R_2]$ (where S = singular points mentioned above) are regular and it follows from Theorem 1. 2-3 that F' is regular in $R - S$. We now examine the singularities more closely. Let b be a singular point on K, i. e., the common endpoint of two curves of the polygon; b is a branch point of multiplicity n and thus $St(b)$ contains more than the two curves of K which meet at b. If $St(b) \cap \text{int } K = 0$ it is clear that the image b' of b will not be a singular point of F'; however, if this is not the case, b' will be, and there will be at least three curves of F' which have the image b' of b as endpoint. Thus by Theorem 2. 2-2 the index $\rho(b')$ of this singularity of F' is $\leqq -\frac{1}{2}$. Hence if we consider the curve $x^2 + y^2 = 4$, whose index is one, we again reach a contradiction, for this curve contains singularities, all of whose indices are negative and thus the requirements of Theorem 2. 2-1 cannot be satisfied. Should the image K' of K contain no singularities of F', we are again reduced to the case proved earlier, since F' will in fact be a branched regular curve family. Whence the theorem is proved.

We shall now describe some other simple configurations which cannot occur in a branched regular curve family. Let us suppose that a cross-section pq crosses a curve C of F at exactly two points t and u between p and q. Denote by $(tu)_1$ the arc of pq between t and u, and by $(tu)_2$ the arc on C. Then $K = (tu)_1 \cup (tu)_2$ is a simple closed curve. We distinguish three possible cases: (i) if p and q are exterior to K, we call this pattern a *bay*, (ii) if one of the points is interior and the other exterior, we call it a *spiral*,

[5] It is suggested that many of the proofs can be followed much more easily by the reader if he makes his own diagrams where none is provided.

and (iii) if both p and q are interior to K, we call it a *harbor* (see Figure 2). From the definition it is seen that the index $\rho(K)$ is $\frac{1}{2}$, 1 and $1\frac{1}{2}$ respectively in the three cases. Since the sum of the indices of the branch points interior to K must be 0 (if there are none) or a negative integer, we have at once:

THEOREM 3. 2-2. *A branched regular curve family filling π can contain neither a bay, nor a spiral, nor a harbor.*

THEOREM 3. 2-3. *The set $L(C+)$ is either empty or contains a single branch point. Similarly $L(C-)$.*

Proof. Suppose $L(C+)$ contained a branch point b and in addition contained points other than b. Let p_n be a sequence of points on C which

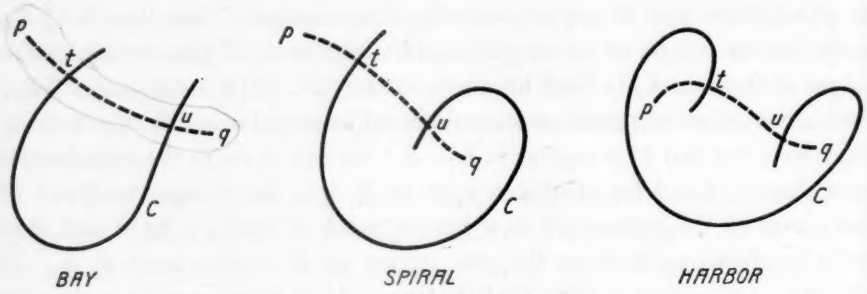

FIGURE 2

approach b, which we assume to have multiplicity n. We consider an admissible neighborhood $U(b)$ with its homeomorphism f carrying U onto the interior of the unit circle. Let F' denote the level curves of $\Re(z^n)$ onto which $F[U]$ is carried homeomorphically by f. Since $L(C+)$ contains points other than b it is clear that $f(C)$ is not one of the radial curves, $\theta = (k/n)\pi$, in F' but must have its image in one or more of the sectors formed by these radial curves, and in fact since $f(p_n) \to f(b)$ the origin, $f(C)$ must cross one of the sectors an infinite number of times. Hence it must be possible to choose points q_n on C near p_n so that q_n approach a regular point q whose image lies on one of the radial lines. Thus we are forced under our assumption to conclude that $L(C+)$ contains at least one regular point q in $F[U(b)]$. Finally, it is clear from the structure of $f(U)$ that this implies that a cross-section through q crosses C an infinite number of times and hence must form at least one bay, spiral, or harbor in so doing. This contradicts the previous theorem; whence it follows that if b is in $L(C+)$, it is the only point in this set.

We remark that this proves in particular that if a curve C is bounded in the positive direction, then $L(C+)$ must be a branch point.

THEOREM 3. 2-4. *An arc pq of $R = \pi - B$ is a cross-section if and only if it meets each chain (or curve) at most once.*

Proof. If pq satisfies the conditions, it is a cross-section by definition. Now suppose pq is a cross-section and that it crosses some chain twice. Let t and u be successive points of crossing and denote by $(tu)_1$ the arc from t to u on pq, and by $(tu)_2$ the arc on the chain. We then have a simple closed curve $K = (tu)_1 \cup (tu)_2$. Interior to K there are only a finite number of branch points, hence by virtue of the preceding theorem we can surely find a curve C through some point r interior to K which leaves K in both directions. Letting r_1 be the first point in the positive direction from r at which C intersects $(tu)_1$, and r_2 the first in the negative direction, then the arcs from r_1 to r_2 on $(tu)_1$ and C must form a bay, a spiral, or a harbor. Since this is a contradiction to Theorem 3. 2-2 the theorem follows.

3.3. Trees. In this section we define an equivalence relation which decomposes the oriented plane, π, into a collection of disjoint closed sets, each of which is a sum of curves of F and points of B, and each of which is a topological tree of a certain type which we define below:

Definition. Let the closed set T of the oriented plane be decomposable into the sum of an at most denumerable collection of subsets C_i, each closed in π, and satisfying the following four conditions:

(1) Each set C_i is the homeomorphic image of either a closed, half-open, or open line segment (whence we will refer to it as a curve).

(2) Each set C_i has at most an endpoint in common with any C_j, $i \neq j$; and if we denote by $St(b)$ the collection of all curves with b as endpoint, then $St(b)$ consists of a finite even number of curves $\geqq 4$.

(3) There is a unique finite chain $c(C_i, C_j) = (C_i, C_{i_2}, \cdots, C_{i_k}, C_j)$ from C_i to C_j for every i, j; i. e., each curve of the chain having an endpoint in common with the preceding curve as in the definitions of **3. 2.**

(4) The sets *open* $St(b)$, consisting of the curves of $St(b)$ without their endpoints opposite b, and *open* C_i, consisting of C_i without its endpoints, are both open sets in T (as a subspace of π).

Then we say that T is a *tree.* (See Figure 3.) Our use of this term is much less general than is usual, but since we consider only this specialized type of tree throughout, there should be no confusion in the use of the term.

The decomposition of any tree T into sets C_i is *unique* quite clearly, except for the numbering, and therefore we may speak without ambiguity of *the* curves of T and *the* endpoints of curves (or, i. e., branch points) of T. Note that a tree is connected and, in fact, arcwise connected by (3) and that by the uniqueness of the chains of (3) there can be no closed curve in T. Condition (4) plus the fact that T is closed in π is equivalent to the following statement: If (p_n) is any sequence of points of T and $p_n \to p \in \pi$, then $p \in T$ and all the points of p_n after some N will lie either on $St(p)$ or on a single curve C_i of T depending on whether p is or is not an endpoint of some curve of T. In the language of combinatorial topology each tree, as described above, is a locally-finite, connected, one-dimensional complex containing no one-cycles. In order to exhibit this, it would be necessary to introduce arbitrarily an infinite number of vertices tending to infinity on each curve of the tree which is homeomorphic to a half-open line segment. Once this is done, the statement is clearly true.

It is clear that in particular any regular curve C of a curve family F is a tree with C being itself the only curve of the decomposition. Now, among the elements of our family F we define the relation *joins* as follows: C is said to *join* C' if and only if there is a finite chain $c(C, C')$ of curves of F from C to C'. If we add to this definition that every curve joins itself, then this is easily shown to be an equivalence relation on the curves of F. We denote by T_C the equivalence class of C, including with each curve its endpoint, i. e., T_C is the set of all curves of F, together with their endpoints, which join C. These equivalence classes are disjoint sets and will be shown below to be trees in the sense of our definition. Thus it will be shown that the curves of a branched regular curve family F fall into classes, each of which is a tree.

THEOREM 3. 3-1. *An arc pq on π is a cross-section of F if and only if it lies entirely in $R = \pi - B$ and has at most one point of intersection with each set T_C.*

Proof. If pq has at most one point in common with each set T_C, since T_C is itself a sum of curves of F with their endpoints, then it will have at most one point in common with each curve of F and hence be a cross-section by definition.

On the other hand, by Theorem 3. 2-4, it is necessary that pq meet any set T_C at most once if it is a cross-section, since if pq met T_C at points r, s, then either C_r, C_s are the same curve or else there is a chain $c(C_r, C_s)$ either of which is impossible by that theorem.

THEOREM 3. 3-2. *Each set T_C of a branched regular curve family F is a tree in the sense of our definition.*

Proof. In the event that C is a regular curve the theorem is trivial since $T_C = C$, as already noted. Now let T_C contain a singular curve, then it follows that it contains only such and at most a countable number, since there are at most a countable number of singular curves in F. Each curve of F in T_C, together with its endpoint(s), will constitute a curve C_i of the decomposition of T_C. Each such set is closed in π, since we include endpoints, and is homeomorphic to either a closed or half-open segment, the latter if the curve extends to infinity in one direction. Thus (1) is satisfied. Condition (2) is, however, also satisfied since each set C_i, being the point set closure of a curve of F, has at most an endpoint in common with any set C_j, $i \neq j$, and, if b is any endpoint, then $St(b)$ contains at least four curves and always an even number, $2n =$ twice the multiplicity of b as a branch point. Likewise (3) is satisfied, i. e., the existence of a chain $c(C, C')$ from $C \subset T_C$ to $C' \subset T_C$ is part of the definition of T_C; and the uniqueness is due to the fact that there can be no polygons of branched curves of F by Theorem 3. 2-1. Finally, we prove simultaneously that condition (4) is satisfied and that T_C is closed as a subset of π. Let p_n be any sequence of points of T_C with a point p of π as limit point. Now if p is a regular point of F, then we take an r-neighborhood $U(p)$ and note that unless every p_n lies on the same curve of F, which is necessarily C_p itself, then we have more than one curve of T_C intersecting $U(p)$ and hence a cross-section through p which must cross T_C more than once, contrary to Theorem 3. 3-1. If p is a branch point, then taking an admissible neighborhood of p, we observe that unless we assume all the points p_n, $n > N$, to lie on $St(p)$, we arrive at the same contradictory conclusion by considering a cross-section from p into one of the sectors of the admissible neighborhood. Thus we conclude that the theorem must be true.

We return to a discussion of a tree T, in the sense of our definition, to study its general structure and its relation to its complementary domain in π. Whether it is or is not a set T_C, i. e. composed of curves of a branched regular curve family or merely abstractly given, is at the moment immaterial. As previously noted, the decomposition of T into curves is unique, and hence we may refer without ambiguity to *the* curves and *the* branch points (or endpoints) of T. Since T is assumed to be imbedded in an oriented plane, a cyclic order is induced on the curves of $St(b)$ [see I]; this allows us to make the following definitions.

We shall call curves C, C' *clockwise adjacent* if they have a common

endpoint b and, in the order induced on $St(b)$ by the orientation of π, moving clockwise the curve C' follows C. In the case of curves of F this means that in each (orientation preserving) homeomorphism of $F[U]$, U an admissible neighborhood of b, onto the level curves of $\Re(z^n)$ in $|z| < 1$, the images of $C \cap U$ and $C' \cap U$ will correspond to radial lines $\theta = (k/n)\pi$, $\theta' = ((k-1)/n)\pi$ respectively. Clearly, given a curve C with endpoint b, there is exactly one curve C' such that C, C' form a clockwise adjacent pair. We call C', C *counterclockwise* adjacent if C, C' are clockwise adjacent. A chain $C_1, \cdots, C_n$ is called an *adjacent chain* if C_i, C_{i+1} are clockwise adjacent for each $i = 1, \cdots, n-1$ or if, for each i, they are counterclockwise adjacent. We shall also consider infinite collections $\{\cdots, C_{-2}, C_{-1}, C_0, C_1, \cdots\}$ of branched curves with the property that for every $k < m$ the subcollection $C_k, \cdots, C_m$ is a chain. Such a collection will be called an *infinite chain* or *half-infinite chain* depending on whether the indices run through both all positive and all negative integers, or whether there is a least negative or a greatest positive integer. An infinite chain will be called adjacent if every subchain is adjacent.

It is convenient at this point to give some attention to a theorem due to Adkisson and MacLane [1] which states that if $\bar{T}$, $\bar{T}'$ are two homeomorphic Peano continua lying on spheres S, S' respectively, then a homeomorphism from $\bar{T}$ to $\bar{T}'$ can be extended to a homeomorphism of S to S' if and only if it preserves the relative sense of every pair of triods of $\bar{T}$. By a triod, $t = [\alpha, \beta, \gamma]$, of $\bar{T}$ is meant any set of three arcs α, β, γ in $\bar{T}$ which have only a single point, called the vertex, in common. A homeomorphism is said to preserve the relative sense of triods of $\bar{T}$ if every two triods t_1, t_2 which have the same sense (i. e., both clockwise or both counterclockwise) on S are carried into two triods t_1', t_2' of $\bar{T}'$ which have the same sense on S'. Let us denote by $\bar{\pi}$ the plane π plus the point ∞ and by $\bar{T}$ the tree T plus the point ∞. Assuming for the moment that the set $\bar{T}$ is a Peano continuum, the theorem above is applicable to our situation. It is a direct consequence of this theorem that, if T, T' are two homeomorphic trees on π and the xy-plane respectively, then any homeomorphism between them may be extended to a homeomorphism of the planes if and only if the relative sense of the curves of $St(b_1)$, $St(b_2)$ is preserved for every pair of branch points b_1, b_2 of T, the only possible vertices for triods being branch points of T and ∞. In order to show that this is a consequence of the theorem, it must only be shown that the relative sense of every pair of triods of $\bar{T}$ is preserved if this is true for every triod of T, i. e. we must consider triods with vertex at ∞. This will be seen to follow from Theorem 6 of the same paper which states that two non-inter-

secting triods $t_1 = [\alpha_1, \beta_1, \gamma_1]$, $t_2 = [\alpha_2, \beta_2, \gamma_2]$ have opposite sense on a sphere S if and only if there exists on S a θ-graph whose vertices are the vertices of t_1 and t_2 and whose three (non-intersecting) arcs contain respectively the legs α_1 and α_2, β_1 and β_2, γ_1 and γ_2. Now it is clear from condition (3) in the definition of a tree (the arcwise connectedness) that given any triod with vertex at ∞, it is possible to find at least one triod with vertex at a branch point of T which is, with it, part of a θ-graph, i. e. a triod with sense opposite that of the given one and with vertex at a branch point. Thus if the sense of triods whose vertices are branch points is always preserved, so also for those whose vertex is at ∞. The conclusion is immediate that we may restate the theorem of Adkisson and MacLane, as we have above, for our own purpose here, if it is true that $\bar{T}$ is a Peano space. But this is clearly true: it is connected by condition (3) of the definition; it is locally connected and locally compact since by condition (4) open C and open $St(b)$ are open sets in T, however these sets have these properties and they cover T. Now T is closed in π by definition, and, on $\bar{\pi}$, ∞ is a limit point of T. By adding this point to T to obtain $\bar{T}$ we clearly obtain a closed set, i. e. $\bar{T}$ is in fact the closure of T in $\bar{\pi}$. Hence $\bar{T}$ is a closed, connected subset of the 2-sphere $\bar{\pi}$. It follows that $\bar{T}$ is compact and hence locally compact. Finally $\bar{T}$ is locally connected since a compact continuum cannot fail to be locally connected at just one point [XI] and, of course, the addition of a single point, ∞, to T does not disturb the local connectedness of T. Thus $\bar{T}$ is a Peano space.

In Section **3. 4** a numbering system for the curves of a tree is described and it is made clear that given any tree it is possible to construct a homeomorphic rectilinear model of it in the xy-plane, i. e. a model consisting of closed and half-open line segments. Moreover, this may even be accomplished so that any one given chain, either finite or infinite, will have its image in the model on a straight line, e. g. on the x-axis. Finally all this may be done preserving the sense in $St(b)$ for every $b \in B$ and thus the homeomorphism of our tree onto its model may be extended to a homeomorphism of π onto the xy-plane. Thus any tree, no matter how badly "twisted," may be straightened out, by a homeomorphism of the entire plane, into a rectilinear tree. This result is established in **3. 4** entirely independently of the remainder of this section, the proof being deferred for convenience only. In order to avoid long and cumbersome proofs the theorems of this section will be established by free use of this fact plus some slight appeal to intuition as far as rectilinear graphs in the plane are concerned.

We now consider relations between a tree T and its complementary domains. In this connection it is convenient to consider a special class of

adjacent chains (of curves of T) which we shall call maximal chains. An adjacent chain of curves of a tree is said to be *maximal* if it is not a subchain of any adjacent chain. It is an immediate consequence of our definitions that a chain of adjacent curves is maximal if and only if (1) it is infinite, or (2) it is half-infinite and its initial (or terminal) curve has only one endpoint, or (3) it is a finite chain and both its initial and terminal curves

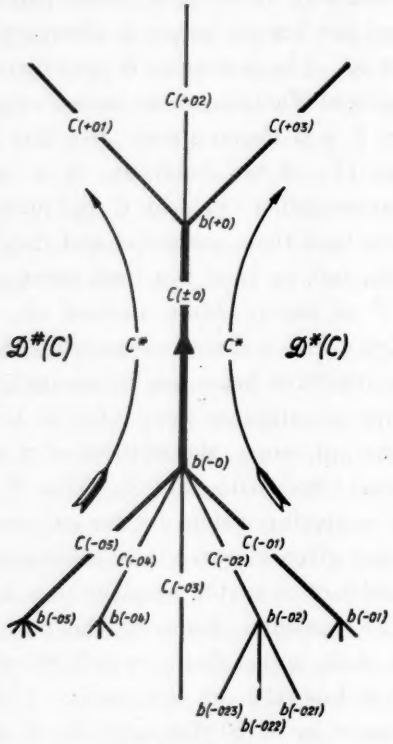

FIGURE 3
RECTILINEAR TREE

have each only one endpoint (a curve of a tree with only one endpoint extends to infinity in the direction opposite to that with the endpoint). Moreover, since a maximal chain is closed as a subset of the tree and a tree is a closed subset of π, so also is every maximal chain a closed subset, and thus is in fact an open curve extending to infinity in each direction, thus dividing the plane into two Jordan domains. (Reference to Figure 3 will clarify this section as well as the following one.)

THEOREM 3. 3-3. *If T is itself a single curve, then it is its only maximal chain. When T contains more than one curve, then* (1) *each curve of T is contained in exactly two maximal chains which intersect only on this curve and* (2) *every branch point of order n is contained in exactly $2n$ maximal chains whose only common point is the branch point itself. Conversely, the intersection of any two maximal chains can be empty, be a single branch point, or, at most a curve of the tree.*

Proof. Let $C_1, \cdots, C_k$ be any clockwise adjacent chain of two or more curves. Now if C_1 has only one endpoint, then there is no curve C' adjacent to C_1 such that $C', C_1, \cdots, C_k$ is a clockwise adjacent chain, since adjacent curves must have an endpoint in common and at a given endpoint there is exactly one curve clockwise adjacent to a given curve. But, if C_1 has two endpoints, then there is exactly one curve C' such that $C', C_1, \cdots, C_k$ is a clockwise adjacent chain. Similar remarks apply to C_k. If neither C_1 nor C_k has more than a single endpoint, then the chain is maximal; in any other case we may extend the chain, one curve at a time added to the initial or final curve, until we arrive at endcurves which have only one endpoint; or, if we do not come to a curve with one endpoint, indefinitely. In any of these cases, the resulting adjacent chain is maximal since it is an open curve extending to ∞ in both directions. Thus every finite adjacent chain which is not alerady maximal can be extended to a unique maximal chain.

If we begin with a single curve C, with at least one endpoint b, then there is one curve clockwise adjacent to C in $St(b)$ and one counterclockwise adjacent. Thus in $St(b)$ we have C, C' and C, C'', unique adjacent chains containing C, one clockwise and one counterclockwise. Hence, C is contained in just exactly two maximal chains, one of which contains C, C', the other C, C''. Similarly, if b is a branch point, there are just $2n$ pairs of adjacent curves in $St(b)$, whence b is contained in $2n$ maximal chains.

Now consider the converse. If two chains intersect, they surely must have a branch point b in common. If this is their only point of intersection in open $St(b)$, then they can intersect at no other point, since the tree is arcwise connected and can contain no closed polygon. If they intersect along two curves of $St(b)$, they must be adjacent curves since the chains are adjacent chains; hence by the preceding remarks on uniqueness they must coincide. This leaves only the possibility that they intersect along a single curve of $St(b)$, and in this case again, since there are no closed curves in the tree, they either have no other intersection or they coincide.

THEOREM 3. 3-4. *Every maximal chain of a tree T divides the plane*

into two domains, whose complete boundary it is; and one of these domains contains no points of T.

Proof. The first part of this theorem is just the Jordan curve theorem. The second part becomes clear when we consider a rectilinear model of the tree. Using the Theorem of Adkisson and MacLane and its consequences as mentioned above, we first map π onto the xy-plane so that the maximal chain becomes the x-axis, every curve of T a chain of line segments, and moreover, so that the orientation is preserved, i. e., so that every clockwise adjacent pair of π will still be clockwise adjacent on the xy-plane, and conversely. Now the contention is that all of the image of T, except what is on the x-axis, will lie in one half-plane, say the upper half-plane. If this is not the case, then there will be a point (u, v) of the upper half-plane and a point (x, y) of the lower half-plane, each in the image T' of T. Then, from $C'(u, v)$, the image-curve containing (u, v), there is a chain to any image-curve on the x-axis, i. e., in the given maximal chain. Let C be the last line segment on the last curve of some such chain to lie in the upper half-plane (more precisely, the endpoint p of C lies on the x-axis, but the rest of the curve lies in the upper half-plane). Similarly, we may choose a line segment C' of T' which lies in the lower half-plane except for one endpoint q. For convenience assume p is to the left of q. Clearly, p and q are branch points on that maximal chain of T' which is the image of our maximal chain in π. Now let C_1, C_2, C_3, C_4 be curves of the maximal chain, i. e., line segments on the x-axis, numbered from left to right, such that p is the common endpoint of the first pair, q of the second. Then it is clear that $[C_1, C, C_2]$ is a triod, with vertex p, of curves in clockwise order, and $[C_3, C, C_4]$ is a triod, with vertex q, of curves in counterclockwise order. But this is impossible since C_1, C_2 and C_3, C_4 are both presumed to be adjacent in the same sense, lying as they do on an adjacent chain. Thus all of T' must lie in either the closed upper half-plane or the closed lower half-plane; whence, the theorem is immediate.

Now let C be a *directed* curve of T, a tree consisting of more than one curve. Then we have seen that C determines exactly two maximal curves which we shall denote by C^* and $C^\#$ with the following convention: as we move along C^* in the direction corresponding to the positive direction on C, then the complementary domain of C^* "to the right" (this can clearly be defined in a topologically invariant manner, by the method used in the proof above) will contain no points of T, and as we move along $C^\#$ in the direction corresponding to the positive direction on C, the complementary domain

"to the left" will contain no points of T. These domains will be denoted $\mathcal{D}^*(C^*)$ and $\mathcal{D}^\#(C^\#)$ respectively, or more simply by $\mathcal{D}^*(C)$ and $\mathcal{D}^\#(C)$ respectively. Now as proved above, C^* is the common boundary of two Jordan domains, and the notation for one of them was given as $\mathcal{D}^*(C^*)$, the other will be denoted by $\mathcal{D}^\#(C^*)$. Similarly, $C^\#$ divides the plane into the domains $\mathcal{D}^\#(C^\#)$ and $\mathcal{D}^*(C^\#)$. When T is just a single curve then C, C^* and $C^\#$ are all the same curve, and $\mathcal{D}^*(C^\#) = 0 = \mathcal{D}^\#(C^*)$. If we reverse the direction on C, we must replace $^\#$ by * throughout.

If we remove *open* C from $C^* \cup C^\#$ we get either two or four half-open arcs extending to infinity from the endpoint(s) of C: two if C has one endpoint, four if it has two. We let $\delta^*(C+)$ denote the arc from the positive endpoint of C lying on C^*, and $\delta^\#(C+)$ the arc from the positive endpoint of C lying on $C^\#$. Similarly, we use the notation $\delta^*(C-)$ and $\delta^\#(C-)$ for the arcs at the other endpoint. We also let $\delta(C+)$ stand for $\delta^*(C+) \cup \delta^\#(C+)$, and $\delta(C-)$ for $\delta^\#(C-) \cup \delta^*(C-)$, and finally, $\delta(C)$ for $\delta(C+) \cup \delta(C-)$. These symbols are clarified by reference to Figure 3.

The collection of all curves C^* and $C^\#$ are then just the maximal chains of T. As already noted above each of these maximal chains bounds two domains, one of which contains no points of T. A converse to this also holds, and, denoting by $\bar{\pi}$ the extended plane and $\bar{T}$ the points of T plus the point at infinity, we have:

THEOREM 3. 3-5. *If T is a tree of π, then $\bar{\pi} - \bar{T}$ consists of an at most countable collection of Jordan domains, each bounded by a simple closed curve in $\bar{T}$ containing the point at infinity. The necessary and sufficient condition that a curve of $\bar{T}$ bound one of those domains is that it be a maximal chain of curves of T.*

COROLLARY 1. *If T is a tree of a regular curve family F, then each complementary domain decomposes into a sum of sets T_C of F.*

COROLLARY 2. *The complementary domains D_n, if infinite in number, tend uniformly to infinity with any sequence p_n, $p_n \in b\,dry\, D_n$, of their boundary points.*

Proof. As we have already noted, π may be mapped on the xy-plane so that the image of T is rectilinear and even so that a given arc (or chain) of T goes onto the x-axis. It is clear then that the complementary domains are Jordan domains. The plane has a countable basis of open sets, whence it follows trivially that the number of complementary domains must be countable also. A boundary curve of a complementary domain must contain

the point at infinity, since T contains no closed curves. Finally, it is clear that π may be mapped onto the xy-plane so that a given complementary domain maps onto the upper half-plane and its boundary onto the x-axis. Thus each such boundary must be a maximal chain.

Corollary 1 follows from the fact that each set T_σ is connected and disjoint from every other such set, and hence must lie in a single complementary domain. If Corollary 2 were not true we would obtain a contradiction to either property (4) of a tree or the fact that T is a closed subset of π.

3.4. A numbering system for the curves of a tree. To facilitate further proofs it will be convenient to establish a system for numbering the curves of a tree of π. The numbering proceeds from an arbitrarily chosen, directed curve C of T, which we shall call the *base curve* of the tree. Using the orientation of the plane together with the existence of a unique chain from the base curve C to each curve of T, we set up a 1-1 correspondence between curves of T and a collection of signed finite sequences, the particular collection depending on both T and C, the sign of the sequences depending only on the direction of C. (See Figure 3.)

To the curve C itself we assign, ambiguously, the sequences ± 0, and we write $C = C(\pm 0)$. If C has a positive endpoint, we denote it by $b(+0)$ and number in *clockwise* order, the curves of $St(b(+0))$ by $[C(+0)]$, $C(+01)$, $C(+02), \cdots, C(0u(+0))$, where $u(+0)$ is defined as the number of curves in $St(b(+0))$ less one, i. e., as twice the multiplicity of that branch point less one. We then denote, if it exists, the endpoint of $C(+0k)$ opposite $b(+0)$ by $b(+0k)$. We follow exactly the same procedure at the other endpoint, if there is one, of $C(\pm 0)$. This endpoint is denoted by $b(-0)$ and the curves of $St(b(-0))$ are numbered, again in *clockwise* order, $[C(-0)]$, $C(-01)$, $C(-02), \cdots, C(-0u(-0))$. If the chain $c(C, C')$ from the base curve C to another curve C' of T contains n curves, we shall say that C' is of *order* n with respect to C. The process above then has numbered every curve of T of order 1 or 2 by *exactly one* finite sequence of one or two elements respectively (except for the ambiguity in the numbering of C itself). Moreover, it assigns a unique sequence to the endpoints of the curve, with the endpoint being numbered with the same number as the curve of lowest order having it as endpoint. Two curves C, C' of the same order will be clockwise adjacent if the final integer of the sequence of C is one less than that of the sequence for C'; and two curves C, C' with C of lower order than C' will be clockwise adjacent if the sequence of C' is that of C with a final integer 1 added to it. Finally,

the chain from the base curve to a curve C' consists of the curves whose numbering sequences are successive "*lower segments*" of the sequence numbering C', i. e., if $\alpha = 0p_2, \cdots, p_n$ numbers C', then $\alpha_1 = 0$, $\alpha_2 = 0p_2$, $\cdots, \alpha_{n-1} = 0p_2 \cdots p_{n-1}$, $\alpha = \alpha_n = 0p_2 \cdots p_{n-1}p_n$ number the curves of the chain from the base curve to C'.

Now if we assume that everything said above is true for every curve of order n, it is very simple to show that it may be extended *in toto* to the curves of order $n+1$; i. e., let C' be any curve of T of order $n+1$. Then C' is the terminal curve of the chain $c(C, C')$ of $n+1$ curves, all of which except C' itself have already received their unique numbering the next-to-last of them by a sequence α of n terms, which sequence also numbers the common endpoint $b(\alpha)$ of this curve and C'. As before, we number the curves in $St(b(\alpha))$ in clockwise order as $[C(\alpha)]$, $C(\alpha, 1)$, $C(\alpha, 2), \cdots, C(\alpha, u(\alpha))$. Here we let α, k denote the sequence whose first n elements are those of α, and whose $(n+1)$-st is k, i. e., we adjoin k to the sequence α. In this process C' will receive a unique numbering, and the statements above will follow through.

With the help of a little new terminology, we will express these facts in a theorem. Given two collections of sequences, A, A^*, we denote by $A \cup A^*$ the collection of all signed sequences obtained by giving those in A positive sign and those in A^* negative. Using this notation we shall call a collection A of finite sequences *allowable* if:

(1) Every sequence has 0 as first element, positive integers for the other elements, and 0 is a sequence of A.

(2) $\alpha, k \in A$ implies $\alpha, k-1 \in A$ if $k \neq 1$, and implies $\alpha \in A$ if $k = 1$.

(3) For each $\alpha \in A$ there is defined an integer $u(\alpha) = 0$ or $\geqq 3$ and odd such that if $u(\alpha) \neq 0$ then $\alpha, 1; \alpha, 2; \cdots; \alpha, u(\alpha)$ are in A but not $\alpha, u(\alpha)+1$; and, if $u(\alpha) = 0$, then there is no sequence of A with α as lower segment, i. e., of the form $\alpha, p_{n+1}p_{n+2} \cdots p_{n+k}$. (Note: If $u(\alpha) = 0$ we call α a *terminal* sequence.)

THEOREM 3. 4-1. *Given a tree T, a curve C of T, and a direction on C, then there exist two unique allowable collections of finite sequences, A, A^*, such that there is a* 1-1 *correspondence between the curves of T and the signed sequences $A \cup A^*$ (except for ± 0 being assigned to C), and such that there is further a* 1-1 *correspondence between the endpoints of the curves of T and those signed sequences which are not terminal of the collection $A \cup A^*$, these correspondences being as described above and having in particular the properties:*

(1) *If $C(\alpha)$ is any curve of T, then $C(\pm 0), C(\alpha_2), \cdots, C(\alpha_{n-1}), C(\alpha)$ is the chain from the base curve to $C(\alpha)$.*

(2) *$C(\alpha, k)$ has the endpoint $b(\alpha)$ in common with the lower order curve $C(\alpha)$ and, if α, k is not a terminal sequence, the endpoint $b(\alpha, k)$ at the opposite end.*

(3) *$C(\alpha)$, $C(\beta)$ of the same order n are clockwise [counterclockwise] adjacent if and only if $\alpha_{n-1} = \beta_{n-1}$ and $\alpha = \alpha_{n-1}, k$; $\beta = \beta_{n-1}, k+1$ $[\beta = \beta_{n-1}, k-1]$. $C(\alpha)$, $C(\beta)$ of different order are clockwise [counterclockwise] adjacent if and only if $\beta = \alpha, 1$ $[\beta = \alpha, u(\alpha)]$.*

It is obvious but tedious to prove that maximal chains, the set $\delta^*(C+)$, $\delta^\#(C+)$ and so on are numbered by sequences with certain characteristic properties. We shall not develop this aspect, but shall state one or two important properties of the numbering below:

THEOREM 3.4-2. *Two trees T on π and T' on π' are homeomorphic under a homeomorphism which may be extended to all of π if and only if we may choose and direct a base curve from each tree and orient the image plane so that the numberings of the two trees are then identical.*

Proof. If the two trees are homeomorphic under such a homeomorphism, then it is clear that for any directed curve C of T as base curve we may choose the homeomorph C' of C as base curve in T', giving it the direction induced by C, and using the orientation in π' induced by the orientation in π, we will get precisely the same numbering for each tree.

On the other hand let T, T' be two trees with identical numberings. We first show that they are homeomorphic. We let $f(\alpha): C(\alpha) \to C'(\alpha)$ be any homeomorphism of $C(\alpha)$ *onto* $C'(\alpha)$ such that $b(\alpha_{n-1})$ maps onto $b'(\alpha_{n-1})$ or, if $\alpha = 0$, so that $b(0+)$ maps onto $b'(0+)$, then $f(\alpha)$ coincides with $f(\alpha_{n-1})$ at $b(\alpha_{n-1})$, the only point where their domains overlap. The map $f: T \to T'$ defined by $f(x) = f(\alpha)x$, for α such that $x \in C(\alpha)$, is 1-1, and is continuous on each of a family of closed sets covering T. Now let x_n be any sequence of points on T such that $x_n \to x \in T$, then by property (4) of trees, for $n \geqq N$, x_n will lie on C_x or $St(x)$, the latter if x is a branch point. From the continuity of f on C_x and $St(x)$ for every $x \in T$, it follows that $f(x_n) \to f(x)$. Hence, since x_n was any sequence and x any point, f is continuous on T. It follows in the same manner that f^{-1} is continuous on T'. Thus f is a homeomorphism from T to T'.

Now in view of the fact that for every branch point b of T, the sense in $St(b)$ must be preserved by f as defined above, by virtue of our earlier slight

modification of the theorem of Adkisson and MacLane, it follows that f can be extended to a homeomorphism on all of π.

We now remark that this theorem makes it possible to construct a rectilinear model of any tree T on the xy-plane, and assures us that there will be a homeomorphism of π onto the xy-plane carrying T onto this model. The model is constructed by considering any numbering $A \cup A^*$ of T, and, using line segments of length $\geqq 1$ as our elements, building up the model piece by piece: We begin with a base element corresponding to the sequence ± 0, add segments corresponding to the 2nd order sequences, 3rd order, etc., each time moving further out from our base segment so that its distance from any n-th order segment approaches infinity with n. In this process it is clearly possible to construct the model so that the image of any one particular chain is a straight line, e. g., the x-axis. (See Figure 3.)

3. 5. r-sets and cross-sections. For an arc pq, lying on an adjacent chain of curves $C_1, \cdots, C_n$, it is possible to get a serviceable analog of the r-neighborhood of an arc on a regular curve (cf. Theorem 1. 2-2). By suitably directing C_1, we have both pq and $C_1, \cdots, C_n$ as arcs on C_1^*. By an *r-set of pq* will be meant any closed set $U \subset \mathcal{D}^*(C_1^*)$ together with a homeomorphism g of U onto the rectangle R_1 of the xy-plane, where $R_1 = \{(x, y) \mid -1 \leqq x \leqq 1, 0 \leqq y \leqq 1\}$, with g having the properties:

(1) g carries $F[U]$ onto the lines $y =$ constant in R_1.

(2) $g^{-1}(\lambda_i)$ are cross-sections, where λ_i, $i = 1, 2$ are, respectively, that part of the lines $x = -1$ and $x = +1$ in R_1.

(3) pq is mapped into the set $\{(x, y) \mid y = 0, -1 < x < 1\}$.

$U(pq)$ contains no branch points except those on C_1^* itself and lies entirely in $\mathcal{D}^*(C_1^*) \cup C_1^*$, which contains no points of T_{C_1} except those on C_1^*. (See Figure 4.) In the event that pq lies on a single curve C_1, then two r-sets which coincide along C_1 but lie on different sides of C_1 will together form an r-neighborhood of C_1. We shall find it convenient to refer to an r-set of a single point p, by which we shall mean one side of a regular neighborhood of p if p is a regular point and one sector of an admissible neighborhood if p is a branch point. We now prove an analog of part of Theorem 1. 2-2.

THEOREM 3. 5-1. *Given any arc pq on an adjacent chain of curves $C_1, \cdots, C_n$, $n > 1$ and any $\epsilon > 0$, there exists an r-set of pq which lies inside the ϵ-neighborhood of pq and in that complementary domain of T whose boundary is the maximal chain containing the given adjacent chain.*

Proof. Let C_1 be directed so that C_1^* contains the chain $C_1, \cdots, C_n$. We leave F unchanged in $\mathfrak{D}^*(C_1^*)$, the complementary domain of T_{C_1} in which U is to lie, but we alter F in the remainder of the plane by mapping the lines $y =$ constant of the lower half-plane, including the x-axis onto $C_1^* \cup \mathfrak{D}^*(C_1^*)$ in such a manner that the x-axis is mapped onto C_1^*. Then

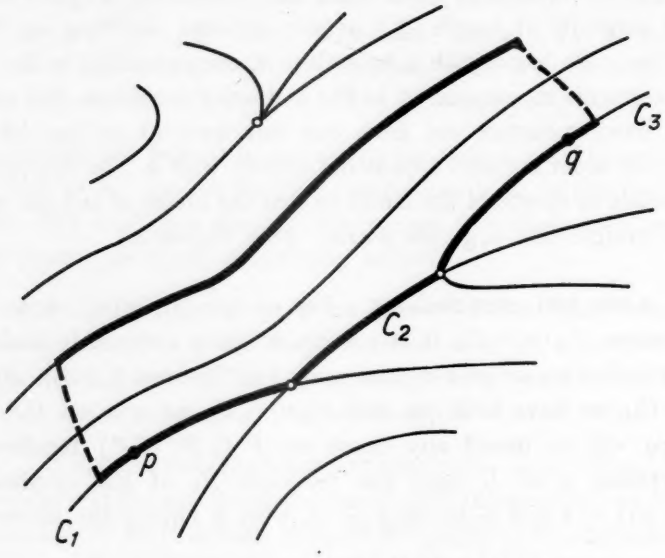

FIGURE 4 AN R-SET

by Theorem 1. 2-3 this new family is regular at all points of C_1^* and agrees with F in $C_1^* \cup \mathfrak{D}^*(C_1^*)$. Hence by Theorem 1. 2-2, there is an arbitrarily small r-neighborhood of pq, call it V. Then $U(pq) = \overline{V \cap [C_1^* \cup \mathfrak{D}^*(C_1^*)]}$ will be our desired r-set.

THEOREM 3. 5-2. *Let a sequence of points q_n on distinct curves C_n approach the point q, where q is a regular point or a branch point, then there is a curve C which may be so directed that q lies on C^* and an infinite subsequence q_m of q_n lies in $\mathfrak{D}^*(C^*)$. Moreover, if p is any point on C^*, then there is a sequence of points p_m in $\mathfrak{D}^*(C^*)$ such that $p_m \to p'$, and such that for $m > M$ the curves C_{p_m} and C_{q_m} are identical.*

Proof. Let T be the tree of F which contains q, then there are at most a finite number of complementary domains of T on whose boundary q lies, and since the q_n lie on distinct curves, there must be a subsequence q_m of these points lying in one of these complementary domains. The maximal chain

which bounds this domain, can, for some suitably directed C be given as $\mathfrak{D}^*(C^*)$. Now let p be any other point on C^*. We may take an r-set $U(pq)$ of pq in $\mathfrak{D}^*(C^*)$ and, if f denotes the homeomorphism $f: U(pq) \to R_1$, then the curves C_m containing q_m map onto lines $y = k_m$ for $m \geqq M$. If $f(p) = (x, 0)$, then the points $p_m = f^{-1}(x, k_m)$ will be the desired sequence.

THEOREM 3. 5-3. *An arc pqr is a cross-section of F if and only if* (1) *it contains no branch points,* (2) *one of the domains $\mathfrak{D}^*(C_q)$, $\mathfrak{D}^\#(C_q)$ contains p, the other r, and* (3) *pq and qr are each cross-sections.*

Proof. We first assume that the arc pqr is a cross-section through q. Then (1) and (3) follow by definition of cross-section. By the Lemma stated in [IV, p. 158] there is an r-neighborhood of q, $V(q)$, such that the image of pqr in R_0 is the y-axis. Every curve crossing $V(q)$ crosses pqr; hence, no curve has more than one line $y =$ constant as image in R_0. The point q itself maps on $(0, 0)$ and C_q on the x-axis; hence, $V - C_q$ splits into two domains, one containing p, whose image in R_0 is $(0, 1)$, and the other r, whose image in R_0 is $(0, -1)$. Moreover, one of these domains lies in $\mathfrak{D}^*(C_q)$ and the other in $\mathfrak{D}^\#(C_q)$, for q is a point on the common boundary of these two domains and hence every neighborhood of q contains points of each domain.

Now, if we assume that pqr is an arc with the properties (1), (2) and (3), we may show that it is a cross-section by showing that it intersects any set T_C at most once. This is clear at once if we remember that a set T_C cannot have points in each of the domains $\mathfrak{D}^*(C_q)$ and $\mathfrak{D}^\#(C_q)$ so that if pqr had more than one point in common with T_C, each such common point would have to lie in the same domain, i. e., both on pq, or both on qr. This is impossible, however, since both of these arcs are cross-sections. It follows that pqr is a cross-section.

The following corollary is immediate:

COROLLARY 1. *If C, C' both intersect a cross-section pq, and each is directed to cross pq in the same direction, then either $\mathfrak{D}^*(C) \supset \mathfrak{D}^*(C')$ or $\mathfrak{D}^\#(C) \subset \mathfrak{D}^\#(C')$.*

COROLLARY 2. *If an arc is such that any point on it is interior to a subarc which is a cross-section, then it is a cross-section.*

Proof. Let pq be such an arc; then we may cover pq with a finite number of r-neighborhoods which overlap. Then, applying the theorem repeatedly a finite number of times gives the result desired.

4. The Family F as the Level Curves of a Continuous Function.

In this section it will be shown that there is a continuous function $f(x, y)$ whose level curves are exactly the family F. The proof of this statement will depend on our ability to remove certain branched curves of F together with their endpoints so as to leave a subset R^* of the plane π, which is open, connected, and simply connected and is such that $F^* = F[R^*]$ is a regular curve family filling R^*. It will then follow from [IV] that there is a continuous function $f^*(x, y)$ defined on R^* and having the family F^* as level curves. Finally, it is shown that $f^*(x, y)$ may be extended to a continuous function on all of the plane with the curves of F as level curves. In this and the next section we will restrict the use of the term *tree* to those sets T_0 containing singular curves, i. e. the term will not be used to include a regular curve.

4. 1. The numbering of the trees of F.

THEOREM 4. 1-1. *If K is any compact subset of π, then there are at most a finite number of distinct trees of F which intersect K on more than one curve of the tree. Moreover, no more than a finite number of curves from any one tree can intersect K.*

Proof. The second part of the theorem is an immediate consequence of the fact that any point p which is a limit of a sequence of points p_n of a tree must be a point of the tree and, in addition to this, for $n \geqq N$, must lie in $St(p)$ if p is a branch point, or on C_p if p is a regular point. If an infinite sequence of curves of a single tree intersected K, we could, by compactness of K, choose a sequence of points on *distinct* curves of this tree such that the sequence would have a limit point, and hence could not conform to the requirements above.

We prove the first part of the theorem by assuming it false and arriving at a contradiction. Let T_i $(i = 1, 2, \cdots)$ be an infinite collection of trees, each intersecting the compact set K on at least the two curves C_i, C_i'. Using the compactness of K we may choose an infinite subsequence from this collection of trees together with points $p_i \in C_i \cap K$ and $q_i \in C_i' \cap K$ on each tree T_i of the subsequence such that $p_i \to p$ and $q_i \to q$. We then throw away all trees except those of our subsequence and renumber. We may assume p is a regular point lying on a curve C and that the p_i all lie in a single complementary domain of T_C, i. e. in $\mathcal{D}^*(C)$. If this is not the case, using Theorem 3. 5-2 we may rechoose p and p_i so that they satisfy these conditions.

This may move them out of K, but this causes no difficulty later. We still have $p_i \in C_i \subset T_i$. In this process, to get the p_i all in a single complementary domain of T_C, it may be necessary to choose a subsequence. If so we throw out the trees T_i containing the points p_i which we remove to get our subsequence, then we renumber. That is, we simultaneously pass to subsequences of p_i and of q_i which correspond, so that we still have $p_i \in C_i$, $q_i \in C_i'$ and and C_i, $C_i' \subset T_i$ after renumbering. This method is followed whenever it is necessary to select a subsequence of either p_i or q_i and renumber. Now we may similarly assume q is a regular point on a curve C' and that the q_i lie in $\mathfrak{D}^*(C')$, or else use Theorem 3. 5-2 as above, and again choose a subsequence. When this process is finished we may be sure that $p \neq q$, for if this were not the case, choosing an r-neighborhood of $p = q$, we would find it crossed by curves C_i, C_i' of the same tree T_i, and this is impossible since it would imply that a cross-section crossed T_i twice.

Now a brief consideration of the image in R_0 of non-intersecting r-neighborhoods $U(p)$ and $U(q)$ and the images of p_i and q_i which lie in these neighborhoods for $i > N$, makes it obvious that we may choose three trees from our sequence of trees: T_{i_1}, T_{i_2}, T_{i_3} which are such that p_{i_1}, p_{i_2}, p_{i_3} lie in $U(p)$ and may be connected in that order by a cross-section γ running from p_{i_1} through p_{i_2} to p_{i_3}; and at the same time q_{i_1}, q_{i_2}, q_{i_3} lie in $U(q)$ and similarly may be connected by a cross-section γ' running from q_{i_1} through q_{i_2} to q_{i_3} (reference to Figure 5 will make the procedure here more easily followed). Then, denoting by c_j the chain $c(C_{i_j}, C_{i_j}')$ in T_{i_j}, $j = 1, 2, 3$, we may define the following three arcs, λ_{i_1}, λ_{i_2}, λ_{i_3} having only their endpoints p_{i_2} and q_{i_2} in common: λ_{i_3} is the arc $p_{i_2}p_{i_1}$ on γ plus the arc $p_{i_1}q_{i_1}$ on c_1 plus the arc $p_{i_1}q_{i_2}$ on γ'. λ_{i_3} is similarly defined with 1 replaced by 3 in the subscripts above, and finally λ_{i_2} is the arc $p_{i_2}q_{i_2}$ on c_2. Two of these arcs, say λ_{j_1}, λ_{j_2}, must form a simple closed curve Γ containing the third λ_{j_3} in its interior. But this is impossible, since each arc contains a branch point; in particular the arc λ_{j_3}, thus enclosed in the interior (in our example) would contain a branch point; and from this branch point issues a chain of curves of T_{j_3}, all distinct from λ_{j_3}, which must leave Γ at some point r. This point r cannot be on T_{j_1} or T_{j_2} since two trees cannot intersect, nor can it be on γ or γ' since then this cross-section would have two points on the same tree, which is ruled out by Theorem 3. 3-1. Hence, we conclude that our initial assumption is impossible and that the theorem must be true.

This theorem will be used to give a method of numbering all non-trivial trees, i. e., trees containing singular curves. We choose any regular point p on π and let K_n designate the circle (with its interior) of center p and

radius n. Now the number of trees cutting K_1 is, of course, denumerable and we number them in any order as $T_{11}, T_{12}, T_{13}, \cdots$ and choose from each a curve C_{11}, C_{12}, C_{13}, respectively, which itself intersects K_1. By the above theorem these choices of curves will be unique for all except a finite number of the trees, and for these C_{1j} is chosen at random from any one of the finite number of curves of the tree cutting K_1. Next we number the trees which

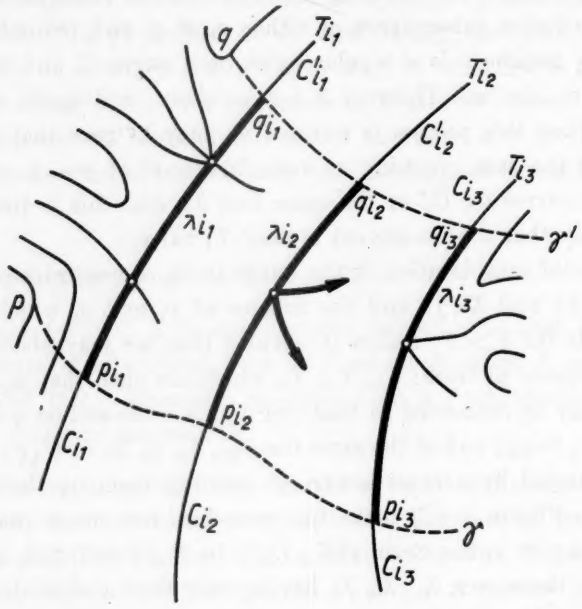

FIGURE 5

intersect K_2 but *not* K_1 as $T_{21}, T_{22}, T_{23}, \cdots$ etc., and let $C_{21}, C_{22}, C_{23}, \cdots$ respectively be curves of these trees which themselves intersect K_2. Proceeding with this process we number the trees cutting K_n but not K_{n-1} as $T_{n1}, T_{n2}, T_{n3}, \cdots$ etc., and choose from each curves $C_{n1}, C_{n2}, \cdots$ cutting K_n. This process will clearly number all the trees of F, and we choose the curves C_{ij} as base curves of the trees, hence determining within each tree T_{ij} a numbering of its curves by sets of finite sequences $A_{ij} \cup A_{ij}^*$ as described in **3.4**. Our method of numbering the trees guarantees that for $m > n$ no tree T_{mj} intersects K_n and, moreover, for each n, there are at most a finite number of curves of the set $\bigcup_{m,j;\, m \leqq n} (T_{mj} - C_{mj})$ which intersect K_n. For future reference we shall call the above method of numbering trees a *standard* numbering of the trees of F.

With these preliminaries we are able to define the curves which we are going to remove from each tree in order to make the multiply connected region $R = \pi - B$ simply connected. Let T be any tree, then it is numbered say as T_{ij} in a standard numbering as described above, and the choice of base curve C_{ij} included in the standard numbering determines a numbering of the curves of $T = T_{ij}$. Let $b(\alpha)$ be any branch point of this tree with sequence $\alpha = e0k_2 \cdot \cdot \cdot k_{n-1}k_n$ with $k_n \neq 1$ (where e denotes the sign $+$ or $-$). Any such $b(\alpha)$ will be called the *initial point* of a cut, and the *cut*, $\lambda(b)$, will consist of all curves $C(\alpha, 1), C(\alpha, 1, 1), C(\alpha, 1, 1, 1), \cdot \cdot \cdot$ and so on *ad infinitum*, or until a terminal sequence $\alpha, 1, 1, \cdot \cdot \cdot, 1$ is reached, i. e., each cut is a chain of adjacent curves extending from $b(\alpha)$ to infinity. We assume endpoints of the curves included, of course, as part of the cut; thus each cut is of the form $\delta^*[C(\alpha)+]$ or $\delta^\#[C(\alpha)-]$, the former if $b(\alpha)$ is the positive endpoint of $C(\alpha)$ and the latter if it is the negative (see heavy lines in Figure 3). Each $\lambda(b)$ is, again, an arc from $b(\alpha)$ to infinity and includes all branch points numbered by sequences of the form $\alpha, 1, 1, \cdot \cdot \cdot, 1$. It is clear from the definition of the cuts and the properties of the numbering that every branch point of the tree is on one and only one cut $\lambda(b)$ and that no two cuts intersect at any point. We denote the collection of all the half-open arcs $\lambda(b)$ on T by $\tilde{T}$, and by $\tilde{\mathfrak{F}}$ the sum of the sets $\tilde{T}$ over all the trees of F. The set $R^* = R - \tilde{\mathfrak{F}}$ contains no branch points and is a union of curves of F. Let F^* denote the family $F[R^*]$ filling R^*.

THEOREM 4. 1-2. *R^* is an open, arcwise connected, and simply connected domain, and F^* is regular in R^*.*

Proof. Let q be any point of R^* and let K_n be the first circle with center at p (in the standard numbering scheme) which contains q in its interior. Now consider how much of K_n is removed when $\tilde{\mathfrak{F}}$ is subtracted from π. None of the base curves C_{ij} is in $\tilde{\mathfrak{F}}$ since none of them is in a set $\lambda(b)$ for these curves are assigned the sequence ± 0 in the numbering, which sequence is not of the form $\alpha, 1, \cdot \cdot \cdot, 1$. And by Theorem 4. 1-1 there can then be at most a finite number of other curves (than base curves) of any T_{ij} in K_n. Hence there is surely an r-neighborhood of q in $K_n - \tilde{\mathfrak{F}}$ and R^* is therefore open and F^* regular.

We wish to show that R^* is arcwise connected. Since every point has an r-neighborhood in R^*, it is clear that R^* is locally-connected. Hence, if it is connected, it is arcwise connected. Now since each cut $\lambda(b)$ is an arc, extending from a point b to infinity, the set $\tilde{\mathfrak{F}} \cup \infty$ on the extended plane $\bar{\pi}$ can clearly be deformed continuously along itself to a single point, the point

at infinity. It follows from Eilenberg [III, Theorem 6, p. 77], that R^* is connected.

Finally, if K is any closed curve in R^* containing a point q of $\tilde{\mathfrak{J}}$, then q lies on a cut $\lambda(b)$ which extends to infinity from b and hence must intersect K, contrary to the assumption that K is in R^*. Thus R^* is simply connected.

THEOREM 4.1-3. *Let $B' \subset B$ be the set of all initial points of cuts $\lambda(b)$, then we may define a collection of disjoint, open sets $\{V_b\}$, one for each $b \in B'$, each containing the corresponding $\lambda(b)$.*

Proof. Referring to the closed circular discs K_n of our standard numbering of the trees of F, we have noted already that only a finite number of the cuts $\lambda(b)$ intersect any K_n. We denote by $B_n' = \{b_j^n\}$ the finite subset of B' whose elements b_j^n are for $j = 1, \cdots, j_n$ those initial points of cuts which intersect K_n but not K_{n-1}. Now, using the normality of π we are able to find disjoint open sets covering the disjoint closed sets $\lambda_1(b_j^1) = \lambda(b_j^1) \bigcap K_1$. We define $V_1(b_j^1)$ as the intersection of the so chosen open sets covering $\lambda_1(b_j^1)$ with the *interior* of K_1. Then we find disjoint open sets covering each of the closed sets $\lambda_2(b_j^i) = \lambda(b_j^i) \bigcap [K_2 - \text{int}\, K_1]$, $i = 1, 2$ and such, moreover, that the open sets covering $\lambda_2(b_j^2)$ do not intersect K_1. Finally, we define $V_2(b_j^i)$, $i = 1, 2$ as the intersections of these open sets with the interior of K_2. Then the sets $V_2(b_j^2)$ and $V_1(b_k^1) \bigcup V_2(b_k^1)$ are non-intersecting open sets lying in the interior of K_2 and covering $\lambda(b_j^i) \bigcap int(K_2)$, for all b_j^i's in B_1' or B_2'. This process is continued indefinitely, covering, for each $\lambda(b)$, its intersection with every K_n and in such a manner that none of the open sets covering one cut has a common point with those covering another cut. Then, given any $b \in B$ it will be in B_n' for some n, hence will be of the form b_i^n, and the cut $\lambda(b_i^n)$ with it as initial point is covered by $V(b_i^n) = \bigcup_{j=n}^{\infty} V_j(b_i^n)$.

THEOREM 4.1-4. *Let F be a branched regular curve family filling the plane π. Then there exists a function $f(p)$ such that:*

(1) *$f(p)$ is defined and continuous for all p in π.*

(2) *for every real number k the locus $f(p) = k$, if not empty, consists of a finite or an at most countably infinite collection of trees (including regular curves) of F.*

(3) *in every neighborhood of any point p in π there are points q for which $f(q) > f(p)$ and points r for which $f(r) < f(p)$.*

Proof. We assume a standard numbering of the non-trivial trees of F and that thus the cuts $\lambda(b)$ and the sets $\tilde{\mathfrak{J}}$ and R^*, etc., are determined. Now this theorem was proved in [IV] by W. Kaplan for curve families regular throughout an open, simply connected domain; thus we may assume that there is a function $f^*(p)$ defined and continuous in R^* and with the properties above. We must show that this function can be extended to a function $f(p)$ with properties (1)-(3) above. The proof, which follows, is divided into three sections, A, B, and C.

(A) First it is necessary to prove that, given any tree T of F, the value of f^* is the same on each curve of $T[R^*]$, i. e., on all curves of T which lie in R^*. Let $C(\pm 0)$ be the base curve of T in the numbering; we shall proceed by induction on the order of the curves of T. If $C(\pm 0)$ has no endpoint, then it is a regular curve, lies entirely in R^* and the result is trivial. Assume it has a positive endpoint $b(+0)$. Then $C(+01)$ is in $\lambda(b(+0))$ and hence not in R^* or $T[R^*]$, but the other curves of $St(b(+0))$ are all in $T[R^*]$. To prove that $f^*(p)$ has the same value on each of these it is only necessary to prove that it has the same value on each pair of adjacent curves among them, for then the value of f^* on $C(+02)$ is the same as that on $C(+03)$ and so on until finally we have the value of $C(+0u(+0))$ the same as that on $C(+0)$. It is quite obvious that this must be so, however, for if C, C' are adjacent curves of $T[R^*]$ and $p \in C$, $q \in C'$, then there is an r-set $U(pq)$ in R^*; and, if $p_n \in U$ is a sequence of points approaching p, then there is a sequence $q_n \in C_{p_n}$ with $q_n \in U$ and q_n approaching q. But, since $q_n \in C_{p_n}$ we have $f^*(q_n) = f^*(p_n)$ and hence $f^*(q) = \lim_{n\to\infty} f^*(q_n) = \lim_{n\to\infty} f^*(p_n) = f^*(p)$. This same procedure actually tells us even more, i. e., that if $C_1, \cdots, C_n$ is any chain of adjacent curves with both $C_1, C_n \subset R^*$, then f^* must have the same value on C_1, C_n.

Now let $C(\alpha, k)$ be a curve of $T[R^*]$ whose sequence is positive and of order $n+1$, and assume that f^* has the same value on each curve of $T[R^*]$ numbered by a positive sequence of order n or less. The sequence α is of the form $\alpha = 0k_2 \cdots k_{n-1}k_n$; and we consider two cases: (1) $k_n \neq 1$ and (2) $k_n = 1$; in either event $k \neq 1$ since $C(\alpha, k)$ is in R^*. In case (1) $b(\alpha)$ is the initial point of a cut, hence $C(\alpha, 1)$ is the only curve of $St(b(\alpha))$ in the cut, and thus the only curve of $St(b(\alpha))$ not in R^*. Moreover, $St(b(\alpha))$ contains the curve $C(\alpha)$ of order n. It follows by precisely the same argument as above that f^* has the same value on each of the curves of $St(b(\alpha))$ in R^* and, in particular on $C(\alpha, k)$, as it has on the n-th order curve $C(\alpha)$ and hence that it has on $C(\pm 0)$. In case (2) both the curves $C(\alpha)$ and

$C(\alpha, 1)$ of $St(b(\alpha))$ are in a cut, i. e. not in R^*. But the curves $C(\alpha_{n-1}, 2)$, $C(\alpha_{n-1}, 1) = C(\alpha)$, and $C(\alpha, u(\alpha))$ form an adjacent chain with the first and last curves in R^*. On the first curve f^* has the same value as on $C(\pm 0)$ since it is of order n, hence it has this value also on the last, $C(\alpha, u(\alpha))$. Now, by going from adjacent curve to adjacent curve, we see that this must be the value of f^* on each curve of $St(b(\alpha))$ in R^* and in particular on $C(\alpha, k)$. This completes the first step in the proof.

(B) Next we define $f(p)$ at every point of π as follows: $f(p) = f^*(p)$ for $p \in R^*$, and $f(p) =$ value of f^* on $T[R^*]$ for $p \in T$. The single-valued function $f(p)$ will then be continuous at each point of R^* (since R^* is an open subset of π and thus the extension cannot affect the continuity of f in that domain), and will have the same value at every point of any tree. Now every point of $\mathfrak{F} = \pi - R^*$ lies on a cut $\lambda(b)$, which in turn lies in a neighborhood $V(\lambda(b))$ not containing points of any other cut. What must be shown is that $f(p)$ is continuous at an arbitrary point q of an arbitrary cut $\lambda(b)$. Now let q_n be any sequence of points approaching the point q of $\lambda(b)$; it must be shown that $f(q_n) \to f(q)$. We shall denote by T the tree containing q; then since $f(p)$ is constant on T, we shall assume that each q_n lies on a distinct curve and none of them is in T. This involves no loss of generality since the result is trivial otherwise. Moreover, we may restrict ourselves to sequences lying in a single complementary domain of T, the reason being that since the number of complementary domains of T with q as a boundary point is finite, any such sequence q_n, with $q_n \notin T$, can be decomposed into a *finite* number of distinct subsequences, containing all the terms of q_n, but no two having a term in common, and each consisting of points from only one complementary domain; and, finally, if for each of these subsequences we have $f(q_{n_i}) \to f(q)$, then $f(q_n) \to f(q)$. Thus we need now to consider only a sequence $q_n \to q$ such that for some $C^* \supset q$, $q_n \in \mathcal{D}^*(C^*)$ for all n. C^* is then in T, and since no cut separates π, there is a curve C' on C^* which is in R^*. Let p be any point of C' and $U(qp)$ an r-set of qp in $\mathcal{D}^*(C^*)$. Then, by Theorem 3. 5-2 there is in U a sequence $p_n \to p$ with $C_{p_n} \equiv C_{q_n}$ and hence $f(p_n) = f(q_n)$. But p is in R^* and $f(p)$ is continuous in R^*, therefore $\lim f(q_n) = \lim f(p_n) = f(p)$. But this is exactly what is needed, for p, q are both on T and hence $f(p) = f(q)$, so f is continuous at q. This establishes property (1) of the theorem.

Property (2) of the theorem is trivial for $f(p)$, since it is satisfied by f^* in R^* and we have added only a denumerable number of curves to the domain of f^* to get the domain of f.

(C) Finally we must prove property (3), i. e., that $f(p)$ has no weak relative extrema. This is clearly equivalent to the following, at least for regular points: if p is a regular point, then f takes a different value on every curve of every r-neighborhod of p, or again equivalently, is monotone on every cross-section. Then, since any arc pq on a curve C has an r-neighborhood, a function satisfies property (3) at every point of a curve or no point of a curve. As to branch points, we can show at once that the condition is satisfied there, for there is always a curve of $St(b)$ in R^*, hence in any neighborhood of b we may find a point q of a curve of R^* and a neighborhood of this point q inside that of b. Now $f(q) = f(b)$ and in this neighborhood of q there will be points q_1, q_2 such that $f(q_1) < f(q) < f(q_2)$, since we are in R^*, where we know f to have property (3). Since q_1, q_2 are in the given neighborhood of b, we have proved our contention for the case of branch points.

Now we wish to show that if f has property (3) on every curve of $St(b)$ except one, C, where b is any branch point, then f has property (3) on C also. Let the curves of $St(b)$ be numbered counterclockwise $C = C_1, C_2, \cdots, C_{2m}$, m being the order of the branch point b. In $U(b)$, an admissible neighborhood, we shall let s_i denote any arc into the sector bounded by C_i, C_{i+1}, such that s_i without b, its endpoint, is a cross-section, e. g., in the image of U on $|z| < 1$ we could take for s_i radii into the respective sectors. Then we indicate by s_i^+ that f increases as we move from b on s_i, by s_i^- that f decreases. Clearly C_i has property (3) if and only if s_{i-1}^+ implies s_i^- and s_{i-1}^- implies s_i^+. Hence if we have s_1^-, then we have by induction s_j^+ for even j, and in particular s_{2m}^+, whence C_1 has property (3), similarly for s_1^+.

Now let the curves of any cut $\lambda(b)$ be numbered $C_1, C_2, \cdots$ beginning with the initial curve and proceeding out from b. C_1 is the only curve of $St(b)$ not in R^*, and hence it must have property (3). If the n-th curve C_n has property (3) then C_{n+1} is the only curve of $St(C_n \cap C_{n+1})$ not having this property since the other curves (than C_n) are in R^*, thus C_{n+1} also must have the desired property. This proves by induction that every curve of every cut has property (3) and hence so has $f(p)$ for all points of π.

A regular curve family F will be said to be *orientable* if its curves may be directed so that every regular point has an r-neighborhood in which the image of each curve in R_0 is directed toward increasing x. Then we have at once:

COROLLARY. *The branched regular curve family F is orientable as a regular curve family in $R = \pi - B$.*

Proof. Exactly as in W. Kaplan [IV, Remark 2, p. 184-5].

NORTHWESTERN UNIVERSITY.

BIBLIOGRAPHY.

[I] V. W. Adkisson and S. MacLane, "Extendnig maps of plane Peano Continua," *Duke Mathematical Journal*, vol. 6 (1940), pp. 216-228.

[II] Y. S. Chin, *Regular families of curves and pseudo-harmonic functions*, Thesis Harvard University, 1947.

[III] S. Eilenberg, "Transformations continues en circonférence et la topologie du plan," *Fundamenta Mathematicae*, vol. 26 (1936), pp. 61-112.

[IV] W. Kaplan, "Regular curve-families filling the plane, I," *Duke Mathematical Journal*, vol. 7 (1940), pp. 154-185.

[V] ———, "Regular curve-families filling the plane, II," *ibid.*, vol. 8 (1941), pp. 11-46.

[VI] ———, "The structure of a curve-family on a surface in the neighborhood of an isolated singularity," *American Journal of Mathematics*, vol. 64 (1942), pp. 1-35.

[VII] ———, "Differentiability of regular curve-families on the sphere," *Lectures in Topology*, Ann Arbor (1941), pp. 299-301.

[VIII] ———, "Topology of level curves of harmonic functions," *Transactions of the American Mathematical Society*, vol. 63, No. 3 (1948), pp. 514-522.

[IX] B. v. Kerékjártó, *Vorlesungen über Topologie*, I, Berlin (1933).

[X] H. Kneser, "Reguläre Kurvenscharen auf den Ringflächen," *Mathematische Annalen*, vol. 91 (1924), pp. 135-154.

[XI] G. T. Whyburn, *Analytic Topology*, American Mathematical Society Colloquium Publications, vol 28, New York (1942).

ON TOTAL NILPOTENT ALGEBRAS.*

By Roy Dubisch and Sam Perlis.

Introduction. A *total nilpotent algebra* of degree n over $\mathfrak{F}$, denoted by $\mathfrak{F}^{(n)}$, is defined to be any isomorphic copy of the algebra of $n \times n$ matrices over $\mathfrak{F}$ with zeros on and above the diagonal. Interest in this class of algebras stems from the following fact (proved in Section 1).

Theorem 1. *The total nilpotent algebras together with their subalgebras constitute the totality of (associative) nilpotent algebras.*

The quantities of $\mathfrak{F}^{(n)}$ may be regarded as linear transformations on the n-dimensional vector space V over $\mathfrak{F}$. Then (Theorem 2) $\mathfrak{F}^{(n)}$ is characterized isomorphically as a maximal algebra of nilpotent linear transformations on $V/\mathfrak{F}$.

The main theme of our investigation of $\mathfrak{F}^{(n)}$ is the explicit determination of all its automorphisms and ideals. The automorphism group $\mathfrak{G}$ is found to have the form $\mathfrak{G} = \mathfrak{D}\mathfrak{M}$, where $\mathfrak{M}$ is a normal subgroup which is "almost all" of $\mathfrak{G}$, and $\mathfrak{D} \cap \mathfrak{M} = I$. Also, $\mathfrak{M} = \mathfrak{J} \times \mathfrak{N}$ where $\mathfrak{J}$ is the group of inner automorphisms (Section 2) and the group $\mathfrak{N}$ consists of all automorphisms carrying every absolute right or left divisor of zero into itself.

After the ideals are determined (Section 7), attention is turned to the characteristic ideals—those which are carried into themselves by every automorphism of $\mathfrak{F}^{(n)}$—and all of these are determined. Certain of the characteristic ideals form a class of algebras which, taken abstractly, are called elementary nilpotent algebras. All total nilpotent algebras are elementary. The final section begins the study of elementary nilpotent algebras.

1. Maximality properties. A total nilpotent algebra $\mathfrak{F}^{(n)}$ has a basis of matric units e_{ij} with $1 \leqq j < i \leqq n$ and is generated by those e_{ij} which lie in $\mathfrak{A} = \mathfrak{F}^{(n)}$ but not [1] in $\mathfrak{A}^2$, namely the quantities

$$e_{21}, e_{32}, \cdots, e_{n,n-1}. \tag{1}$$

Thus, $e_{ij} = e_{i,i-1}e_{i-1,i-2} \cdots e_{j+1,j}$. The product of the $n-1$ quantities (1)

* Received February 14, 1950; presented to the American Mathematical Society, August 31, 1949.

[1] The reader can verify that e_{ij} lies in $\mathfrak{A}^k$ if and only if $i - j \geqq k$. Hence e_{ij} lies in $\mathfrak{A}^k$ but not in $\mathfrak{A}^{k+1}$ if and only if $i - j = k$.

in reverse order is $e_{n1} \neq 0$, but every product of n factors chosen from the list (1) is zero. It follows that $\mathfrak{A}^{n-1} > \mathfrak{A}^n = 0$, that is, $\mathfrak{F}^{(n)}$ is nilpotent of index n. To complete the proof of Theorem 1, let $\mathfrak{B}$ be any nilpotent algebra of order $n-1$ and let $\mathfrak{B}_1$ denote the algebra $(1, \mathfrak{B})$ obtained by adjunction of a unity to $\mathfrak{B}$. The regular representation of $\mathfrak{B}_1$ contains a subalgebra $\mathfrak{B}_0$ isomorphic to $\mathfrak{B}$. If in obtaining the regular representation we use a normalized basis [2] of $\mathfrak{B}$, we find [3] that the matrices in $\mathfrak{B}_0$ belong to $\mathfrak{F}^{(n)}$.

THEOREM 2. *For every nilpotent sub-algebra $\mathfrak{N}$ of the total matric algebra $\mathfrak{F}_n$ of degree n over $\mathfrak{F}$ there is an automorphism T of $\mathfrak{F}_n/\mathfrak{F}$ carrying $\mathfrak{N}$ into $\mathfrak{N}^T \leqq \mathfrak{F}^{(n)}$.*

Proof. The algebra $\mathfrak{N}$ defines an algebra $\mathfrak{L}$, isomorphic to $\mathfrak{N}$, of linear transformations on the n-dimensional vector space V over $\mathfrak{F}$, and $\mathfrak{N}$ represents $\mathfrak{L}$ relative to the basis $e_1, \cdots, e_n$ where e_i is the vector whose i-th coordinate is 1 and whose other coordinates are 0. Thus if $N = (\nu_{ij}) \in \mathfrak{N}$, $N \to L$ where $e_i{}^L = \Sigma \nu_{ij} e_j$. The images v^L $(v \in V, L \in \mathfrak{L}^i)$ span a vector space $V\mathfrak{L}^i$ and $V\ (= V\mathfrak{L}^0) > V\mathfrak{L} > \cdots > V\mathfrak{L}^{\nu-1} > V\mathfrak{L}^\nu = 0$ where ν is the index of nilpotency of $\mathfrak{L}$. For, the fact that $\mathfrak{L}^{\nu-1} \neq 0$ implies that $V\mathfrak{L}^{\nu-1} \neq 0$. If $V\mathfrak{L}^i = V\mathfrak{L}^{i+1}$ for $i < \nu$, it follows that $V\mathfrak{L}^i = V\mathfrak{L}^{i+1} = \cdots = V\mathfrak{L}^\nu = 0$, a contradiction. Now we choose representatives of the classes constituting any basis of $V\mathfrak{L}^{\nu-1} - V\mathfrak{L}^\nu$, then of $V\mathfrak{L}^{\nu-2} - V\mathfrak{L}^{\nu-1}$, etc., and finally of $V - V\mathfrak{L}$, and number these quantities $v_1, v_2, \cdots, v_n$ in the order of choice. The quantities $v_1, \cdots, v_n$ so obtained form a basis of V. Relative to this basis each linear transformation T is represented by a matrix (α_{ij}) defined by the equation $v_i{}^T = \Sigma \alpha_{ij} v_j$ and if T is in $\mathfrak{L}$, the construction shows that $\alpha_{ij} = 0$ for $j \geqq i$. Thus $\mathfrak{F}^{(n)}$ contains the algebra $\mathfrak{N}_0$ of the (α_{ij}). Since $\mathfrak{N}_0$ and $\mathfrak{N}$ represent $\mathfrak{L}$ relative to different bases of V, it is known that $\mathfrak{N}_0 = P^{-1}\mathfrak{N}P$ where P is a non-singular matrix of $\mathfrak{F}_n$.

This result characterizes $\mathfrak{F}^{(n)}$ as a maximal algebra of nilpotent [4] linear transformations on the n-dimensional vector space over $\mathfrak{F}$. There is no such algebra of larger dimension and any other of the same dimension is isomorphic to $\mathfrak{F}^{(n)}$.

---

[2] A basis $u_1, \cdots, u_{n-1}$ such that each product $u_i u_j$ is a linear combination of those u_k with k greater than i and j. Cf. L. E. Dickson, *Algebras and their Arithmetics*, Chicago, 1923, pp. 175 f.

[3] Theorem 1 is clearly a corollary of Theorem 2. The separate proof of Theorem 1 is given both for its simplicity and for the notations and preliminary properties which it introduces.

[4] For an algebra $\mathfrak{A}$, nilpotency of all the quantities of $\mathfrak{A}$ is equivalent to the nilpotency of $\mathfrak{A}$.

2. Monic and inner automorphisms. If $\mathfrak{A}$ is an arbitrary (associative) ring, an automorphism T of $\mathfrak{A}$ is called "monic" in case $x^T - x$ lies in $\mathfrak{A}^{r+1}$ whenever x lies in $\mathfrak{A}^r$. This concept is trivial if $\mathfrak{A}$ possesses a unity but otherwise the concept is significant.

Let T and U be monic and consider the equation $x^{TU} - x = (x^{TU} - x^T) + (x^T - x)$. Since x^T lies in $\mathfrak{A}^r$ if x does, both quantities in parentheses lie in $\mathfrak{A}^{r+1}$, so that TU is monic. Also

$$x^T - x = y \in \mathfrak{A}^{r+1}, \qquad x^{T^{-1}} - x = -y^{T^{-1}} \in \mathfrak{A}^{r+1},$$

so that T^{-1} is monic. Hence the monic automorphisms form a group, $\mathfrak{M}$. This group is always a normal subgroup of the automorphism group $\mathfrak{G}$ of $\mathfrak{A}$. For, let us take $M \in \mathfrak{M}$, $G \in \mathfrak{G}$, $x \in \mathfrak{A}^r$. Then $x^G \in \mathfrak{A}^r$, $x^{GM} - x^G = y \in \mathfrak{A}^{r+1}$,

$$x^{GMG^{-1}} - x = y^{G^{-1}} \in \mathfrak{A}^{r+1}.$$

Thus GMG^{-1} is monic.

THEOREM 3. *The monic automorphisms of any associative ring*[5] $\mathfrak{A}$ *form a normal subgroup* $\mathfrak{M}$ *of the automorphism group of* $\mathfrak{A}$. *If* $\mathfrak{A}$ *is nilpotent, the group* $\mathfrak{M}$ *is nilpotent.*

To prove the latter statement, let $\mathfrak{M}_i$ be the subset of automorphisms $M \in \mathfrak{M}$ such that $x^M - x = x_i \in \mathfrak{A}^i$. It is easily verified that $\mathfrak{M}_i$ is a subgroup and that $\mathfrak{M} = \mathfrak{M}_2 \geqq \mathfrak{M}_3 \geqq \cdots \geqq \mathfrak{M}_\nu = I$, where $\mathfrak{A}^{\nu-1} > \mathfrak{A}^\nu = 0$. Then $\mathfrak{M}$ is nilpotent[6] if each commutator $(\mathfrak{M}, \mathfrak{M}_i)$ is contained in $\mathfrak{M}_{i+1}$. Using the convention that a subscript j on a quantity of $\mathfrak{A}$ indicates that the quantity belongs to $\mathfrak{A}^j$, we have for each $L \in \mathfrak{M}_i$ and each $M \in \mathfrak{M}$:

$$x^{MLM^{-1}} = (x + x_2)^{LM^{-1}} = (x + x_2 + x_i)^{M^{-1}} = x + x_i^{M^{-1}} = x + x_i + x_{i+1}$$
$$= (x + x_2 + x_i) + (x_{i+1} - x_2),$$

$$x^{MLM^{-1}L^{-1}} = (x + x_2) + (x_{i+1} - x_2)^{L^{-1}} = x + x_2 + x_{i+1} - x_2 + y_{i+1} = x + z_{i+1}.$$

This completes the proof.

Inner automorphisms may be defined in any ring $\mathfrak{A}$. An element $a \in \mathfrak{A}$ is called quasi-regular if there is a $b \in \mathfrak{A}$ such that $a + b + ab = a + b + ba = 0$. Then an "inner" automorphism J of $\mathfrak{A}$ is defined by the formula

$$x^J = x + bx + xa + bxa. \tag{2}$$

Proof that (2) gives an automorphism is obtained by considering a ring $\mathfrak{B}$

[5] This theorem and its proof remain valid if $\mathfrak{A}$ is an algebra over $\mathfrak{F}$ and the automorphisms considered are over $\mathfrak{F}$.

[6] H. Zassenhaus, *The Theory of Groups*, New York, 1949, pp. 125 f.

containing $\mathfrak{A}$ and a unity element 1. Then $1 + a$ is regular with $1 + b$ as its inverse in $\mathfrak{B}$. The inner automorphism $y \to (1 + b)y(1 + a) = y^J$ carries each $x \in \mathfrak{A}$ into a quantity $x^J \in \mathfrak{A}$ given by (2). These images x^J vary over every quantity $u \in \mathfrak{A}$ since $x = (1 + a)u(1 + b)$ is in $\mathfrak{A}$ and $x^J = u$. Thus J induces an automorphism in $\mathfrak{A}$.

The totality of automorphisms $J = J_a$ given by (2) as a varies over the quasi-regular elements of $\mathfrak{A}$ form a group $\mathfrak{J}$ and, from (2), it is evident that each J is monic, $\mathfrak{J} \leqq \mathfrak{M}$. If G is any automorphism of $\mathfrak{A}$,

$$x^{G^{-1}JG} = x + b^G x + x a^G + b^G x a^G$$

which is the same as (2) with a^G and b^G in place of a and b. But a^G is quasi-regular with b^G as its quasi-inverse, so that $G^{-1}JG$ is again an inner automorphism. In summary: The inner automorphisms of any associative ring $\mathfrak{A}$ are all monic and form a normal subgroup of the automorphism group of $\mathfrak{A}$.

3. Certain characteristic ideals. A subalgebra $\mathfrak{C}$ of an algebra $\mathfrak{A}$ over $\mathfrak{F}$ is called "characteristic" if every automorphism T of $\mathfrak{A}$ over $\mathfrak{F}$ induces an automorphism in $\mathfrak{C}$. The powers $\mathfrak{A}^k$ are characteristic ideals. The knowledge that x lies in a particular characteristic ideal of $\mathfrak{A}$ is a clue to the possible quantities x^T. We shall begin to track down the automorphisms T of $\mathfrak{A} = \mathfrak{F}^{(n)}$ by finding for each matric unit e_{ij} in $\mathfrak{A}$ certain characteristic ideals containing e_{ij}.

Define $\mathfrak{C}_l$ to be the totality of matrices in $\mathfrak{F}^{(n)}$ whose columns beyond the l-th are zero. Then $\mathfrak{C}_l$ is an ideal of $\mathfrak{F}^{(n)}$. Likewise, an ideal is given by the set $\mathfrak{R}_k$ of all matrices in $\mathfrak{F}^{(n)}$ whose first $k - 1$ rows are zero. We shall see [7] that $\mathfrak{C}_l$ and $\mathfrak{R}_k$ are characteristic.

The algebra $\mathfrak{A}^l$ consists (see footnote 1) of all $n \times n$ matrices with zeros everywhere except for an area in the lower left hand corner which may be described as a right triangle with both legs having length $n - l$ units. The elements in this area are completely arbitrary. It follows quickly that $\mathfrak{C}_l$ is precisely the totality of quantities $c \in \mathfrak{F}^{(n)} = \mathfrak{A}$ such that $c\mathfrak{A}^l = 0$, that is, $\mathfrak{C}_l$ is the left annihilator of $\mathfrak{A}^l$ in $\mathfrak{A}$. Then for each $c \in \mathfrak{C}_l$, $x \in \mathfrak{A}^l$, and automorphism T of $\mathfrak{A}$, we have $cx = 0 = (cx)^T = c^T x^T$, and since x^T varies over all of $\mathfrak{A}^l$ as x does, $c^T\mathfrak{A}^l = 0$, so $c^T \in \mathfrak{C}_l$. Thus $\mathfrak{C}_l$ is a characteristic ideal of $\mathfrak{A}$.

The ideal $\mathfrak{R}_k$ can be described as the totality of quantities $r \in \mathfrak{A} = \mathfrak{F}^{(n)}$ such that $\mathfrak{C}_{k-1}r = 0$ (the right annihilator of $\mathfrak{C}_{k-1}$ in $\mathfrak{A}$). Since $\mathfrak{C}_{k-1}$ is

[7] We are indebted to the referee for simplifications in the proofs of Theorems 4 and 7.

characteristic, the final argument above may be paraphrased to show that $\mathfrak{R}_k$ is characteristic.

THEOREM 4. *The ideals $\mathfrak{C}_l$ and $\mathfrak{R}_k$ of $\mathfrak{F}^{(n)}$ are characteristic. For each automorphism T of $\mathfrak{F}^{(n)}$ and each matric unit $e_{kl} \in \mathfrak{F}^{(n)}$, $e_{kl}{}^T$ lies in $\mathfrak{C}_l \cap \mathfrak{R}_k$.*

Thus, in particular,

$$e_{k,k-1}{}^T = \sum_{i \geqq k}^{j \leqq k} \beta_{ij}{}^{(k)} e_{ij} \qquad (\beta_{ij}{}^{(k)} \in \mathfrak{F}). \tag{3}$$

The "corner" coefficient

$$\beta_{k,k-1}{}^{(k)} = \alpha_k$$

cannot be zero for otherwise $e_{k,k-1}{}^T$ would lie in the square of $\mathfrak{F}^{(n)}$ whence so would $e_{k,k-1}$, contrary to fact. Moreover, if T is monic, each $\alpha_k = 1$; the converse can be verified easily from the fact that $\mathfrak{F}^{(n)}$ is generated by the quantities $e_{k,k-1}$.

4. Diagonal automorphisms. The set $\mathfrak{T}_n$ of all matrices in $\mathfrak{F}_n$ with zeros above the diagonal is an algebra whose radical is $\mathfrak{F}^{(n)}$. The radical of any algebra is characteristic and we shall be interested in the automorphisms induced in $\mathfrak{F}^{(n)}$ by certain inner automorphisms of $\mathfrak{T}_n$.

Each diagonal non-singular matrix, $d = \Sigma \delta_i e_{ii}$ $(\delta_i \in \mathfrak{F})$, determines an automorphism $x^D = dxd^{-1}$ of $\mathfrak{T}_n$, and the automorphism which D induces in $\mathfrak{F}^{(n)}$ is called a "diagonal" automorphism of $\mathfrak{F}^{(n)}$. We have

$$e_{k,k-1}{}^D = \alpha_k e_{k,k-1}, \tag{4}$$

and

$$e_{kl}{}^D = \alpha_{kl} e_{kl}, \qquad \alpha_{kl} = \alpha_k \alpha_{k-1} \cdots \alpha_{l+1}, \tag{5}$$

where the α_k are non-zero scalars. Conversely, arbitrary non-zero scalars $\alpha_2, \cdots, \alpha_n$ and equations (4) and (5) determine a diagonal automorphism given by $d = e_{11} + \alpha_2 e_{22} + \cdots + (\alpha_2 \cdots \alpha_n) e_{nn}$. Each diagonal D is obtained precisely once in the form $x^D = dxd^{-1}$ where $d = e_{11} + \delta_2 e_{22} + \cdots + \delta_n e_{nn}$, and, consequently, the group $\mathfrak{D}$ of all diagonal automorphisms of $\mathfrak{F}^{(n)}$ is isomorphic to the multiplicative group of non-singular diagonal matrices in $\mathfrak{F}_{n-1}$.

Next we find that from each automorphism T a unique diagonal factor D may be removed so that $D^{-1}T$ is monic.

THEOREM 5. *The automorphism group of $\mathfrak{F}^{(n)}$ has the structure $\mathfrak{G} = \mathfrak{M}\mathfrak{D} = \mathfrak{D}\mathfrak{M}$ where $\mathfrak{D}$ is the group of diagonal automorphisms, $\mathfrak{M}$ the group of monic automorphisms, and $\mathfrak{D} \cap \mathfrak{M} = I$.*

In (3) and the remarks after (3) it appeared that certain "corner coefficients" $\alpha_k = \beta_{k,k-1}{}^{(k)}$ cannot be zero. By multiplying equations like (3) for the quantities $e_{k,k-1}{}^T, \cdots, e_{l+1,l}{}^T$ we find that

$$e_{kl}{}^T = \sum_{i \geqq k}^{j \leqq l} \alpha_{ij} e_{ij}$$

with corner coefficient $\alpha_{kl} = \alpha_k \alpha_{k-1} \cdots \alpha_{l+1}$. The quantities α_k, α_{kl} thus fulfill the conditions in (5) so that the correspondence $e_{kl} \to \alpha_{kl} e_{kl}$ generates a diagonal automorphism D. Then $D^{-1}T = T_1$ and $TD^{-1} = T_2$ are automorphisms whose corner coefficients are all unity and this implies that T_1 and T_2 are monic.

That $\mathfrak{D} \cap \mathfrak{M} = I$ is evident from (4) and (5). Thus in any expression $T = DM = M_1 D_1$ the factors are unique and, actually, $D_1 = D$ since $\mathfrak{M}$ is normal in $\mathfrak{G}$.

The idea of inner automorphisms of Section **2** may be applied to $\mathfrak{A} = \mathfrak{F}^{(n)}$. Here every $a \in \mathfrak{A}$ is quasi-regular with quasi-inverse $b = -a + a^2 - \cdots + (-1)^t a^t$ $(a^{t+1} = 0)$. With the aid of the unity in $\mathfrak{T}_n$ we may write $g = 1 + a = 1 + \Sigma \alpha_{ij} e_{ij} = g_2 g_3 \cdots g_n$, where

$$g_i = 1 + \sum_{j=1}^{i-1} \alpha_{ij} e_{ij} = 1 + a_i. \tag{6}$$

Then $g_i{}^{-1} = 1 - a_i$ and

$$x^J = g^{-1} x g = g_n{}^{-1} \cdots g_2{}^{-1} x g_2 \cdots g_n, \qquad J = J_a = J_{a_2} \cdots J_{a_n}. \tag{7}$$

These properties will be needed soon.

5. Nil automorphisms. A quantity $u \in \mathfrak{A}$ is an absolute right (left) divisor of zero if $xu = 0$ $(ux = 0)$ for every $x \in \mathfrak{A}$. For $\mathfrak{A} = \mathfrak{F}^{(n)}$ these are the quantities $e_{n1}, e_{n2}, \cdots, e_{n,n-1}$ $(e_{21}, e_{31}, \cdots, e_{n1})$ and their linear combinations.

An automorphism N of $\mathfrak{A}$ is called "nil" if $u^N = u$ for every absolute right or left divisor of zero. It is clear that the totality of nil automorphisms of any ring form a group $\mathfrak{N}$ which is actually a normal subgroup of the full automorphism group.

THEOREM 6. *All nil automorphisms of $\mathfrak{F}^{(n)}$ are generated by*

$$e_{k,k-1}{}^N = e_{k,k-1} + \gamma_k e_{n1} \qquad (k = 2, \cdots, n) \tag{8}$$

where $\gamma_2 = \gamma_n = 0$ and the remaining γ_k are arbitrary scalars.

From (8) it follows that $e_{kj}{}^N = e_{kj}$ for $j < k - 1$, whence $e_{k1}{}^N = e_{k1}$ and $e_{nj}{}^N = e_{nj}$ so that N is nil. Conversely, let N be nil. By Theorem 5 $N = DM$,

$$e_{k,k-1}{}^D = \alpha_k e_{k,k-1}, \qquad e_{k,k-1}{}^{DM} = \alpha_k e_{k,k-1} + c_k$$

with $c_k \in \mathfrak{A}^2$. Then $\alpha_2 = 1$ and $c_2 = 0$. If $k > 2$,

$$e_{k1} = e_{k1}{}^N = (e_{k,k-1}e_{k-1,1})^N = e_{k,k-1}{}^{DM}e_{k-1,1} = (\alpha_k e_{k,k-1} + c_k)e_{k-1,1}$$
$$= \alpha_k e_{k1} + c_k e_{k-1,1}$$

so $\alpha_k = 1$ and $c_k e_{k-1,1} = 0$. By the first of these facts, $D = I$ and N is monic. By the second the $(k-1)$-st column of the matrix c_k is 0. For $1 < j < k-1$,

$$0 = (e_{k,k-1}e_{j1})^N = (e_{k,k-1} + c_k)e_{j1} = c_k e_{j1}$$

so that the j-th column of c_k is 0 for all $j > 1$. For $n > j > k$,

$$0 = (e_{nj}e_{k,k-1})^N = e_{nj}{}^N e_{k,k-1}{}^N = e_{nj}(e_{k,k-1} + c_k) = e_{nj}c_k$$

so the j-th row of c_k is 0, $j = k+1, \cdots, n-1$. Also

$$(e_{nk}e_{k,k-1})^N = e_{n,k-1}{}^N = e_{n,k-1} = e_{nk}(e_{k,k-1} + c_k) = e_{n,k-1} + e_{nk}c_k,\ e_{nk}c_k = 0$$

so that the k-th row of c_k is also 0, $c_k = \gamma_k e_{n1}$, and this gives (8).

THEOREM 7. *The group of monic automorphisms of $\mathfrak{F}^{(n)}$ has the structure $\mathfrak{M} = \mathfrak{N} \times \mathfrak{I}$, where $\mathfrak{N}$ is the group of nil automorphisms and $\mathfrak{I}$ the group of inner automorphisms.*

To prove that $\mathfrak{M} = \mathfrak{N}\mathfrak{I}$ it suffices to find for each $M \in \mathfrak{M}$ an automorphism $J \in \mathfrak{I}$ such that $MJ = N \in \mathfrak{N}$, whence $M = NJ^{-1}$. We have

$$e_{k1}{}^M = e_{k1} + \sum_{p>k} \alpha_{pk} e_{p1} \qquad (k = 2, \cdots, n-1).$$

The scalars α_{pk} are thus defined for $n \geqq p > k > 1$, and

$$e_{k1}{}^M = e_{k1} + \sum_{p>q>1} \alpha_{pq}e_{pq}e_{k1} = (1+a)e_{k1} = (1+a)e_{k1}(1+a)^{-1},$$
$$a = \sum_{p>q>1} \alpha_{pq}e_{pq}.$$

The inner automorphism $x^{J_1} = g^{-1}xg$, $g = 1 + a$, has the property that $M_1 = MJ_1$ fixes each element e_{k1}. Since M_1 is monic, $e_{nk}{}^{M_1} = e_{nk} + \sum_{q<k} \beta_{kq}e_{nq}$, and for all p such that $1 < p < k$ we have $e_{nk}e_{p1} = 0 = (e_{nk}e_{p1})^{M_1} = e_{nk}{}^{M_1}e_{p1}$ $= \sum_{q<k} \beta_{kq}e_{nq}e_{p1} = \beta_{kp}e_{n1}$. It follows that the coefficients β_{kp} are zero except possibly for $\beta_{k1} = \beta_k$ $(k = 2, \cdots, n-1)$ so

$$e_{nk}{}^{M_1} = e_{nk} + \beta_k e_{n1} = e_{nk} + e_{nk}\sum_{j=2}^{n-1} \beta_j e_{j1}$$
$$= e_{nk}(1+b), \qquad b = \sum_{j=2}^{n-1} \beta_j e_{j1}.$$

Then the inner automorphism $x^{J_2} = (1+b)x(1+b)^{-1}$ has the properties

$$e_{nk}{}^{M_1J_2} = e_{nk}, \qquad (k = 1, \cdots, n-1),$$

$$e_{k1}{}^{M_1J_2} = e_{k1}{}^{J_2} = (1+b)e_{k1}(1+c) = e_{k1}.$$

Thus $M_1J_2 = MJ_1J_2$ is nil. This proves $\mathfrak{M} = \mathfrak{N}\mathfrak{J}$.

Since $\mathfrak{N}$ and $\mathfrak{J}$ are normal subgroups of $\mathfrak{M}$, the product $\mathfrak{N}\mathfrak{J}$ is the direct product $\mathfrak{N} \times \mathfrak{J}$ if and only if the intersection $\mathfrak{N} \cap \mathfrak{J}$ is the identity. Let J be in this intersection: $x^J = (1+a)^{-1}x(1+a)$. Since J also lies in $\mathfrak{N}$, $e_{k1}{}^J = e_{k1}$, $(1+a)e_{k1} = e_{k1}(1+a)$ whence $ae_{k1} = e_{k1}a = 0$. The matrix a thus has all columns zero except possibly its first column. Similarly, J fixes each e_{nk}, $0 = ae_{nk} = e_{nk}a$, and a has only zero rows except perhaps for row n. Thus $a = \alpha e_{n1}$, $\alpha \in \mathfrak{F}$,

$$(1+a)^{-1} = 1 - \alpha e_{n1}, \qquad x^J = (1-\alpha e_{n1})x(1+\alpha e_{n1}) = x.$$

This proves $J = I$ and completes the proof of the theorem.

6. The automorphism group of $\mathfrak{T}_n$. Each diagonal automorphism $x^D = d^{-1}xd$ and each inner automorphism $x^J = (1+a)^{-1}x(1+a)$ of $\mathfrak{F}^{(n)}$ may be extended to $\mathfrak{T}_n$ over $\mathfrak{F}$ simply by letting x vary over $\mathfrak{T}_n$. The theorem below implies that the automorphisms DJ of $\mathfrak{F}^{(n)}$ are the only ones extendable to $\mathfrak{T}_n$ and it determines all extensions of these.

THEOREM 8. *The group $\mathfrak{E}$ of automorphisms of $\mathfrak{T}_n/\mathfrak{F}$ inducing the identity on $\mathfrak{F}^{(n)}$ is given by $x^E = g^{-1}xg = (1-\alpha e_{n1})x(1+\alpha e_{n1})$ where α is an arbitrary scalar. All automorphisms of $\mathfrak{T}_n/\mathfrak{F}$ are inner and the group they induce in $\mathfrak{F}^{(n)}$ is $\mathfrak{D}\mathfrak{J}$.*

Since each automorphism T of $\mathfrak{T}_n/\mathfrak{F}$ maps the e_{ii} into a set of mutually orthogonal idempotents whose sum is 1, it can be shown that

$$e_{kk}{}^T = e_{j_kj_k} + b_k \tag{9}$$

where b_k is in $\mathfrak{F}^{(n)}$ and $k \to j_k$ is a permutation of $1, 2, \cdots, n$; and

$$e_{j_kj_k}b_k + b_ke_{j_kj_k} + b_k{}^2 = b_k, \tag{10}$$

$$\sum_{k=1}^{n} b_k = 0. \tag{11}$$

To show that the nil automorphisms N are not extendable to $\mathfrak{T}_n$ unless $N = I$, let T induce N on $\mathfrak{A} = \mathfrak{F}^{(n)}$. Then by Theorem 6,

$$(e_{kk}e_{k,k-1})^T = e_{k,k-1} + \alpha_k e_{n1} = (e_{j_kj_k} + b_k)(e_{k,k-1} + \alpha_k e_{n1})$$

$$= e_{j_kj_k}e_{k,k-1} + c_k \qquad (c_k \in \mathfrak{A}^2, k = 2, \cdots, n),$$

so $e_{j_k j_k} e_{k,k-1} = e_{k,k-1}$, $j_k = k$ $(k = 2, \cdots, n)$ whence also $j_1 = 1$ and

$$e_{kk}{}^T = e_{kk} + b_k \qquad (k = 1, \cdots, n).$$

Also

$$(e_{k,k-1} e_{k-1,k-1})^T = e_{k,k-1} + \alpha_k e_{n1} = (e_{k,k-1} + \alpha_k e_{n1})(e_{k-1,k-1} + b_{k-1})$$
$$= e_{k,k-1} + (e_{k,k-1} b_{k-1} + \delta \alpha_k e_{n1}) \qquad (k = 2, \cdots, n),$$

where $\delta = 0$ unless $k - 1 = 1$. Then the quantity in parenthesis must equal $\alpha_k e_{n1}$. If T is not the identity on $\mathfrak{F}^{(n)}$, there is a value of k such that $n > k > 2$ and $\alpha_k \neq 0$ (Theorem 6). Choosing this k we find $\delta = 0$, $e_{k,k-1} b_{k-1} = \alpha_k e_{n1} \neq 0$, which is impossible since the n-th row of the left side is 0. This proves that nil automorphisms $\neq I$ are not extendable, so the only extendable automorphisms are those of $\mathfrak{DF}$.

If T is any automorphism of $\mathfrak{T}_n$, it induces an automorphism DJ in $\mathfrak{F}^{(n)}$, and the latter is clearly extendable to an inner automorphism S of $\mathfrak{T}_n$ such that $TS^{-1} = E$ is the identity on $\mathfrak{F}^{(n)}$. It remains only to show that each automorphism E inducing the identity in $\mathfrak{F}^{(n)}$ is inner and obeys the formula of Theorem 8. Then $T = ES$ will be inner.

For $i < k = 2, \cdots, n$ we have, by (9), $(e_{kk} e_{ki})^E = e_{ki} = (e_{j_k j_k} + b_k) e_{ki}$, whence $j_k = k$ and $b_k \in \mathfrak{F}^{(n)}$ has k-th column zero. Then $b_k e_{kk} = 0$ and (10) yields

$$b_k = e_{kk} b_k + b_k{}^2 \qquad (k = 2, \cdots, n). \tag{12}$$

On squaring both sides we find

$$b_k{}^2 = e_{kk} b_k{}^3 + b_k{}^4 \tag{13}$$

which implies that $b_k{}^2$ lies in $\mathfrak{A}^3$, $\mathfrak{A} = \mathfrak{F}^{(n)}$. When this fact is employed on the right side of (13) it turns out that $b_k{}^2$ lies in $\mathfrak{A}^4$ and, similarily, $b_k{}^2$ lies in $\mathfrak{A}^t$, $t = 3, 4, \cdots$. But then $b_k{}^2 = 0$ and (12) reduces to $b_k = e_{kk} b_k$, that is, b_k is zero except possibly for its k-th row $(k = 2, \cdots, n)$. A parallel proof shows that b_k is 0 except possibly for its k-th column $(k = 1, \cdots, n - 1)$. It follows that $b_2 = \cdots = b_{n-1} = 0$ and, from (11), $b_1 = -b_n$. But b_1 has only a first column, b_n only an n-th row, so $b_n = \alpha e_{n1}$, $b_1 = -\alpha e_{n1}$. Thus $e_{ij}{}^E = e_{ij}$ except for $e_{11}{}^E = e_{11} - \alpha e_{n1}$, $e_{nn}{}^E = e_{nn} + \alpha e_{n1}$, and this is the automorphism E given in Theorem 8.

7. Ideals. A set of r matric units $f_i = e_{k_i l_i} \in \mathfrak{F}^{(n)}$ will be called a "set of corners" if $r < n$ and $k_1 < k_2 < \cdots < k_r$, $l_1 < l_2 < \cdots < l_r$. Each corner may be represented by the point in an $n \times n$ matrix where its only non-zero coordinate occurs, and these points determine a "staircase" as

follows. Regard the matrix as a square array of n^2 points, and draw a line in row k_1 extending from the left boundary of the square to the corner $f_1 = e_{k_1 l_1}$. Continue the line from this corner down to row k_2, then to the right until it reaches f_2. In general, if $i < r$, the line goes from corner f_i down to row k_{i+1}, then right to corner f_{i+1}. From f_r the line simply goes down to row n; or the line stops at f_r if f_r lies in row n. The open polygon so obtained has the appearance of a staircase in which each corner, except possibly f_r, is at the top of a riser. We shall refer to this open polygon as the "staircase" determined by the given set of corners. A matric unit e_{ij} is an "interior quantity" for the set of corners if its representative point lies below the staircase or on it but not at a corner.

THEOREM 9. *Let $f_1, \cdots, f_r$ be any set of corners and $u_1 \cdots, u_s$ be linear combinations of these corners with scalar coefficients such that each f_i occurs in at least one u_j with non-zero coefficient. Then $u_1, \cdots, u_s$ and the interior quantities for the set of corners span an ideal of $\mathfrak{F}^{(n)}$ and, conversely, every ideal is obtained in this way.*

Proof that the space spanned is an ideal is routine. In proving that every ideal $\mathfrak{B}$ is obtained in the manner described we first find the corners. Among all the e_{ij} occurring in the expressions of the quantities of $\mathfrak{B}$ in terms of matric units, let $i = k_1$ denote the minimum value of i. Among all the $e_{k_1 j}$ occurring let $j = l_1$ be maximum. This determines f_1. Among all e_{ij} occurring with $i > k_1$ and $j > l_1$ (if any) let $i = k_2$ denote a minimum value and, for this i, let $j = l_2$ be maximum. This determines f_2 and a continuation yields all the corners $f_1, \cdots, f_r$, $r < n$. The non-zero coordinates of all matrices in $\mathfrak{B}$ occur on or under the staircase and it will be shown that $\mathfrak{B}$ contains all the interior quantities.

For each f_i there is a quantity u_i in $\mathfrak{B}$ containing f_i with coefficient unity. Let $f_i = f = e_{kl}$ and $u_i = u$. If $f = e_{n1}$, this must be the only corner and $\mathfrak{B}$ has the form required, so we assume that e_{n1} is not a corner. If $l = 1$, $\mathfrak{B}$ contains $e_{jk}u = e_{j1}$ $(j = k+1, \cdots, n)$ and if $k = n$, $\mathfrak{B}$ contains $ue_{1m} = e_{nm}$ $(m = 1, \cdots, l-1)$. If $l > 1$ and $k < n$, $\mathfrak{B}$ contains every $e_{jk}ue_{1m} = e_{jm}$ $(j = k+1, \cdots, n; m = 1, \cdots, l-1)$, that is, every e_{jm} lying below and to the left of $f = e_{kl}$. Since this is true for every f not at an end of the staircase, $\mathfrak{B}$ contains every interior quantity under the staircase. If any of the u_i involve interior quantities under the staircase, these may be subtracted off, so it will be assumed henceforth that the u_i are linear combinations of the corners and possibly other matric units lying on the staircase.

If u and $f = e_{kl}$ are as above, $\mathfrak{B}$ contains $e_{jk}u = e_{jl} + v$ where v is a

linear combination of interior quantities, hence in $\mathfrak{B}$, so $\mathfrak{B}$ contains e_{jl} for all $j > k$. Likewise, $\mathfrak{B}$ contains all e_{km}, $m < l$, that is, $\mathfrak{B}$ contains every staircase quantity to the left of or below any corner. By use of these quantities the u_i may be altered so that they involve only corners.

If $x \in \mathfrak{B}$, subtraction of interior quantities leaves a quantity $y = \Sigma \alpha_i f_i \in \mathfrak{B}$. If y is not a linear combination of the u_i, it will be taken as an additional quantity u. This process stops after a finite number of repetitions since all the quantities y and u lie in the r-dimensional vector space spanned by the corners. The u_i now at hand, together with the interior quantities, span $\mathfrak{B}$. This completes the proof.

8. Characteristic ideals. A subalgebra $\mathfrak{B}$ of $\mathfrak{A}$ is called "characteristic" if $\mathfrak{B}^T = \mathfrak{B}$ for every automorphism T of $\mathfrak{A}$. If $\mathfrak{B}$ is an ideal of $\mathfrak{F}^{(n)}$ and M is a monic automorphism, M carries each e_{ij} in $\mathfrak{B}$ or u_i (Theorem 9) in $\mathfrak{B}$ into the same quantity plus an interior quantity of $\mathfrak{B}$, so $\mathfrak{B}^M = \mathfrak{B}$. Therefore $\mathfrak{B}$ is characteristic if and only if $\mathfrak{B}^D = \mathfrak{B}$ for each diagonal automorphism D. This condition is satisfied if the corners are contained in $\mathfrak{B}$ since D merely multiplies each matric unit by a scalar. By the next theorem this is essentially the only case in which $\mathfrak{B}$ is a characteristic ideal.

THEOREM 10. *If $\mathfrak{F}$ contains only two elements, every ideal of $\mathfrak{F}^{(n)}$ is characteristic. Otherwise, an ideal $\mathfrak{B}$ of $\mathfrak{F}^{(n)}$ is characteristic if and only if $\mathfrak{B}$ contains its corners.*

The proof of the first statement is the observation that every diagonal automorphism is the identity.

LEMMA 1. *If $f_1, \cdots, f_r$ are a set of corners and $\phi_1, \cdots, \phi_r$ are arbitrary non-zero scalars, there is a diagonal automorphism D such that $f_i^D = \phi_i f_i$ $(i = 1, \cdots, r)$.*

First we construct $D = D_a$ such that

$$f_a^D = \phi_a f_a, \qquad f_i^D = f_i \qquad (i \neq a), \tag{14}$$

by finding a suitable diagonal matrix $d = \delta_1 e_{11} + \cdots + \delta_n e_{nn}$ to use for $x^D = dxd^{-1}$. Let $f_a = f = e_{kl}$ so $l < k$ and $f^D = \delta_k \delta_l^{-1} f$. Then the definitions $\delta_k = \phi_a$, $\delta_j = 1$ for $j < k$, yield (14) for $i = 1, \cdots, a$. Let $f_j = e_{k_j l_j}$ for $a \leqq j < r$ and assume that $\delta_1, \cdots, \delta_{k_j}$ have been defined so that (14) holds for $i = 1, \cdots, j$ and that for $h > k_j$, δ_h is still undefined. If $f_{j+1} = e_{pq}$ then $f_{j+1}^D = \delta_p \delta_q^{-1} f_{j+1}$ and δ_p is still undefined. If δ_q is among the δ's already defined, we take $\delta_p = \delta_q$, and $\delta_h = 1$ for $k_j < h < p$. If δ_q is undefined,

we take $\delta_h = 1$ for $k_j < h \leqq p$. Then $f_{j+1}{}^D = f_{j+1}$, whence the induction is completed. The D of the lemma is given by $D = D_1 D_2 \cdot \cdot \cdot D_r$.

Now let $\mathfrak{F}$ contain more than two elements. The sufficiency has been noted above, so it remains to show that a characteristic ideal $\mathfrak{B}$ contains its corners. The u_i of Theorem 9 may be taken to be linearly independent, so $s \leqq r$ and the equations

$$u_i = \sum_{j=1}^{r} \beta_{ij} f_j \qquad (i = 1, \cdot \cdot \cdot, s)$$

define an $s \times r$ matrix $\mathbf{B} = (\beta_{ij})$ of rank s. Since $\mathfrak{B}$ is characteristic it contains

$$u_i{}^D = \sum_{j=1}^{r} \beta_{ij} \phi_j f_j \qquad (i = 1, \cdot \cdot \cdot, s) \tag{15}$$

for all non-zero $\phi_j \varepsilon \mathfrak{F}$ (Lemma 1). But if $u_i{}^D$ is in $\mathfrak{B}$ it must be of the form $\Sigma \xi_{ih} u_h$, which implies that

$$\sum_j \beta_{ij} \phi_j f_j = \sum_{h,j} \xi_{ih} \beta_{hj} f_j \qquad (i = 1, \cdot \cdot \cdot, s)$$

or, in matrix form,

$$\mathbf{B}\Phi = \Xi \mathbf{B}, \tag{16}$$

where $\Xi = (\xi_{ik})$ is an $s \times s$ matrix depending on the diagonal matrix Φ whose diagonal elements are $\phi_1, \cdot \cdot \cdot, \phi_r$. Since $\mathbf{B}$ is $s \times r$ of rank s it has a right inverse Δ, that is, an $r \times s$ matrix Δ such that $\mathbf{B}\Delta = I_s$, the s-rowed identity. Therefore $\Xi = \mathbf{B}\Phi\Delta$ and (16) implies that

$$\mathbf{B}\Phi\mathbf{H} = 0, \qquad \mathbf{H} = (\eta_{ij}) = I_r - \Delta\mathbf{B}. \tag{17}$$

Note that $\mathbf{H}$ is independent of Φ and that (17) is true for all choices of Φ. Taking $\phi_j = 1$ for $j \neq a$ we find that the element in the (i, l) place in (17) is

$$\beta_{ia} \phi_a \eta_{al} + \sum_{j \neq a} \beta_{ij} \eta_{jl} = 0 \tag{18}$$

for all choices of $\phi_a \neq 0$. Thus (18) is a linear equation in ϕ_a with more than one solution (note the hypothesis on F!)—a contradiction unless every $\beta_{ia}\eta_{al} = 0$. For each value of a there is a value of i such that $\beta_{ia} \neq 0$ (each f_a occurs in some u_i). Therefore, every $\eta_{al} = 0$, $\mathbf{H} = 0$, and $\Delta\mathbf{B} = I_r$. Since $\Delta\mathbf{B}$ cannot have higher rank than $\mathbf{B}$, the matrix $\mathbf{B}$ must have rank r. Hence $s = r$. Since the space $(f_1, \cdot \cdot \cdot, f_r)$ of dimension r has a subspace $(u_1, \cdot \cdot \cdot, u_r)$ of the same dimension, the two coincide, whence all the f_i lie in $\mathfrak{B}$.

COROLLARY. *If $\mathfrak{F}$ is not of characteristic two, the number of characteristic ideals of $\mathfrak{F}^{(n)}$ is* $\frac{1}{n}\binom{2n}{n-1}$.

The proof of this fact, which was kindly supplied us by G. Szekeres, will be omitted here.

9. Elementary nilpotent algebras. An ideal with a single corner $f = e_{kl}$ is the intersection $\mathfrak{D}_{kl} = \mathfrak{R}_k \cap \mathfrak{C}_l$ of the characteristic ideals $\mathfrak{R}_k$ and $\mathfrak{C}_l$ defined in Section **3** and, therefore, every characteristic ideal is a join of such intersections $\mathfrak{D}_{kl}$. This is one of several facts lending interest to the ideals $\mathfrak{C}_l = \mathfrak{C}_l{}^{(n)}$ as a class of abstract algebras. It is unnecessary to study also the algebras $\mathfrak{R}_k$ since $k = n - l + 1$ for an appropriate l and $\mathfrak{R}_{n-l+1} \cong \mathfrak{C}_l$ under the correspondence $e_{n-j+1,n-i+1} \to e_{ij}$. The algebras $\mathfrak{D}_{kl}$ are of no structural interest since they are zero algebras.

Any algebra over $\mathfrak{F}$ isomorphic to $\mathfrak{C}_l{}^{(n)}$ will be called an elementary nilpotent algebra. Note that $\mathfrak{F}^{(n)} = \mathfrak{C}_{n-1}{}^{(n)}$ so that $\mathfrak{F}^{(n)}$ is elementary. By direct computation, the index of $\mathfrak{C}_l{}^{(n)}$ is found to be $l + 1$.

THEOREM 11. *Two elementary nilpotent algebras are isomorphic if and only if they have the same values of l and of n. The invariant l is equal to $\nu - 1$ where ν is the index of nilpotency.*

The sufficiency is evident. Suppose, conversely, that $\mathfrak{C}_k{}^{(m)} \cong \mathfrak{C}_l{}^{(n)}$. The common index of nilpotency of these algebras is $k + 1 = l + 1$, whence $k = l$. The common order is $l[n - (l+1)/2] = l[m - (l+1)/2]$, so $n = m$.

THEOREM 12. $\mathfrak{C}_l{}^{(n)}$ *is indecomposable unless* $l = 1$.

Since $\mathfrak{C}_l{}^{(n)}$ is a zero algebra, it is the direct sum of $n - 1$ ideals of order one. Now let l be greater than 1 and suppose that $\mathfrak{C}_l{}^{(n)} = \mathfrak{A} \oplus \mathfrak{B}$. Then one of $\mathfrak{A}$ and $\mathfrak{B}$, say $\mathfrak{A}$, contains an element of the form

$$a = e_{21} + \sum_{i \neq 2} \alpha_{ij} e_{ij} \qquad (\alpha_{ij} \in \mathfrak{F}) \tag{19}$$

whence $\mathfrak{A}$ contains every $e_{j2}a = e_{j1}$ $(j = 3, \cdots, n)$. If $\mathfrak{B}$ contained an element (19) it would also contain e_{n1}, contrary to the fact that $\mathfrak{A} \cap \mathfrak{B} = 0$. Hence every non-zero $b \in \mathfrak{B}$ has the form $b = \Sigma \beta_{ij} e_{ij}$ with $\beta_{21} = 0$. Then if all $\beta_{ij} = 0$ for $j > 1$, b would lie in $\mathfrak{A}$, $b = 0$, a contradiction implying that some $\beta_{pq} \neq 0$ for $q > 1$. Then $\mathfrak{B}$ contains $be_{q1} = \Sigma \beta_{iq} e_{i1} \neq 0$ and be_{q1} is a linear combination of $e_{31}, \cdots, e_{n1}$. Hence be_{q1} is in $\mathfrak{A} \cap \mathfrak{B}$. This contradiction shows that $\mathfrak{B}$ has no quantities $b \neq 0$.

The center of any nilpotent algebra $\mathfrak{N}$ of index ν contains the highest non-zero power $\mathfrak{N}^{\nu-1}$ of the algebra and the center is "minimal" if it equals this power. The center of $\mathfrak{C}_l^{(n)}$ (hence, also, of $\mathfrak{F}^{(n)}$) is minimal. A more inclusive result is given now.

THEOREM 13. *Let* $\mathfrak{A} = \mathfrak{C}_l^{(n)}$. *Then the difference algebras* $\mathfrak{A} - \mathfrak{A}^h$ $(h = 2, 3, \cdots, l+1)$ *have minimal centers* $\mathfrak{A}^{h-1} - \mathfrak{A}^h$.

First it should be noted that $\mathfrak{A}$ is generated by the matric units on its "right border," that is, by $e_{21}, e_{32}, \cdots, e_{l+1,l}, e_{l+2,l}, \cdots, e_{nl}$. Also the successive powers of $\mathfrak{A}$ are obtained by removing successive lines along the right border.

The theorem is true for $h = 2$ since $\mathfrak{A} - \mathfrak{A}^2$ is a zero algebrra. Assuming the result for $\mathfrak{A} - \mathfrak{A}^h$ $(h < l+1)$ we now let $cx - xc$ lie in $\mathfrak{A}^{h+1}$ for every x of $\mathfrak{A}$ and wish to prove c in $\mathfrak{A}^h$. Since $cx - xc$ is always in $\mathfrak{A}^h$, c must be in $\mathfrak{A}^{h-1}$, $c = c_0 + d$ where c_0 is in $\mathfrak{A}^h$ and

$$d = \sum_{j=1}^{l-h+1} \alpha_{h-1+j,j} e_{h-1+j,j} + \sum_{k=l+1}^{n} \alpha_{k,l-h+2} e_{k,l-h+2}. \tag{20}$$

It suffices to prove $d = 0$. From $d = c - c_0$ it follows that $dx - xd \in \mathfrak{A}^{h+1}$ for every x. Hence $\mathfrak{A}^{h+1}$ contains the quantities $a_i = de_{i,i-1} - e_{i,i-1}d$ $(i = 2, \cdots, l)$. The non-zero elements of the matrix $b_i = de_{i,i-1}$ lie in the $(i-1)$-st column and those of $c_i = e_{i,i-1}d$ lie to the left of this column. Then the fact that a_i lies in $\mathfrak{A}^{h+1}$ implies that b_i and c_i also lie in $\mathfrak{A}^{h+1}$. But $b_i = \alpha_{h-1+i,i} e_{h-1+i,i-1}$ $(i = 2, \cdots, l-h+1)$, which is in $\mathfrak{A}^{h+1}$ only if every $\alpha_{h-1+i,i} = 0$. That the first summation in (20) is zero now follows from $c_{h+1} = e_{h+1,h}d = \alpha_{hl} e_{h+1,1}$ which is in $\mathfrak{A}^{h+1}$ only if $\alpha_{hl} = 0$. Since $\mathfrak{A}^{h+1}$ contains

$$b_{l-h+2} = \sum_{k=l+1}^{n} \alpha_{k,l-h+2} e_{k,l-h+1}$$

it follows that every coefficient $\alpha_{k,l-h+2}$ is 0 and thus that $d = 0$.

FRESNO STATE COLLEGE AND PURDUE UNIVERSITY.

A THEORY OF SUBINVARIANT LIE ALGEBRAS.*

By EUGENE SCHENKMAN.**

A theory of subinvariant subgroups has been developed by Wielandt [7][1] who proved by way of application of the theory that the tower of automorphisms of a finite group with no center ends finitely. A similar study of subinvariant Lie algebras is presented here.

A subalgebra of a Lie algebra L is said to be subinvariant in L if it is a member of a sequence of subalgebras ending with L, each an ideal in the following. Some of the elementary properties of subinvariant subalgebras are given in Section 2. It is shown there that the intersection of two subinvariant subalgebras is subinvariant and that the property of subinvariance is preserved under homomorphisms; namely, if A is subinvariant in L, then $A\nu = \bar{A}$ is subinvariant in $\bar{L} = L\nu$, where ν is a homomorphism. Furthermore, if $\bar{A}$ is subinvariant in $\bar{L}$ and if A is the complete inverse image of $\bar{A}$ under ν, then A is subinvariant in L. We then show that if A and B are subinvariant in L, the composition factors from B to $A \cap B$ occur among those from L to A.

In Section 3 we prove that a subinvariant subalgebra is the sum of an ideal and a nilpotent algebra. This is done by proving that if A is subinvariant in L then the intersection $A^\omega = \bigcap_{k=1}^{\infty} A^k$ of all the power algebras of A is an ideal of L, and then by showing that an arbitrary algebra A is the sum of the ideal A^ω and a nilpotent algebra. These results are true for arbitrary base fields.

In Sections 4 through 7 we require that the base field be of characteristic zero. Section 4 is devoted to showing that the radical and nil radical of a subinvariant subalgebra are the intersections, respectively, of the subalgebra with the radical and nil radical of the containing algebra. The proof given is based on Cartan's decomposition theorem for a semisimple Lie algebra by means of which is is shown that the radical and nil radical are characteristic ideals.

In Section 5 it is shown that the algebra $\{A, B\}$, generated by two

* Received May 25, 1950. This paper is part of a dissertation presented for the degree of Doctor of Philosophy in Yale University.

** The author was an Atomic Energy Commission Predoctoral Fellow in the Physical Sciences of the National Research Council during the course of this research.

[1] Numbers in brackets refer to the bibliography.

subinvariant subalgebras A and B of L, is subinvariant in L; and that the different composition factors in a composition series from $\{A, B\}$ to A occur, not counting multiplicities of one dimensional factors, among those in a composition series from B to $A \cap B$. To show this last fact we prove that Lie algebra L can be expressed as the sum of its radical and the ideal $L^{(\omega)} = \bigcap_{k=1}^{\infty} L^{(k)}$, the intersection of all the derived algebras of L.

The problem of finding the smallest subinvariant subalgebra containing a given element of a Lie algebra L is considered in Section 6. It is shown that this subalgebra is the sum of an ideal and the one-dimensional algebra determined by the element itself.

In Section 7 we give several results relating an arbitrary subalgebra V of L to the greatest subinvariant subalgebra of L contained in V, and to the smallest subinvariant subalgebra of L containing V.

The last two section are concerned with an application of the theory to arbitrary Lie algebras. We show first in Section 8 that if A is a subinvariant subalgebra of L, and if $A^{\omega} = \bigcap_{k=1}^{\infty} A^k$ has a centralizer not contained in A^{ω}, then A has a non-zero centralizer in L. With this result we prove in Section 9 that the tower of derivation algebras, $L = D_0, D_1, \cdots, D_n, \cdots$ of a Lie algebra with no center over an arbitrary field ends finitely with $\dim D_n \leq d + c$, where c is the dimension of the center of $L^{\omega} = \bigcap_{k=1}^{\infty} L^k$, and d is the dimension of the derivation algebra of L^{ω}. As an immediate consequence of our tower theorem we have Chevalley's result [2] for characteristic zero that there exists an n such that D_n has only inner derivations.

It is of interest to contrast the theory of subinvariant Lie algebras with the theory of subinvariant groups. Almost without exception Wielandt's results [7] have analogues for Lie algebras; but there are important differences between the methods of the two theories. In the first place one has the concept of conjugate subgroups which is frequently used throughout the theory of subinvariant subgroups. This concept is available for nilpotent subinvariant subalgebras over fields of characteristic zero or for arbitrary subalgebras over the fields of real or complex numbers, but not in general. On the other hand the useful fact that if A is a subinvariant subalgebra of L, then $A^{\omega} = \bigcap_{k=1}^{\infty} A^k$ is an ideal of L, does not have an analogue for groups. Neither is there any analogue in group theory to Cartan's decomposition theorem for a semisimple Lie algebra which enables one to give the form of the minimal subinvariant subalgebra determined by a single element.

Before developing the theory I wish to express my sincere appreciation to Professor Nathan Jacobson for suggesting this research and for his constant

interest and stimulating discussions during the course of its progress. I should also like to thank Professors Shizuo Kakutani and C. E. Rickart, and Mr. Charles Curtis for reading portions of the manuscript.

1. Preliminaries. We recall that L is a Lie algebra over a field Φ if L is a vector space over Φ in which the composition is defined such that for $l, m, n \in L$, $\alpha \in \Phi$

$$[l, l] = 0 \tag{1}$$

and

$$[[l, m], n] + [[m, n], l] + [[n, l], m] = 0 \tag{2}$$

in addition to the usual relations

$$[l, m + n] = [l, m] + [l, n]; \qquad [l + m, n] = [l, n] + [m, n]$$

and

$$\alpha[l, m] = [\alpha l, m] = [l, \alpha m].$$

We are concerned only with finite-dimensional algebras.

If A is a subspace of L and if $[a, l] \in A$ for every $a \in A$, $l \in L$, then A is an *ideal* or *invariant subalgebra* of L. It will be convenient to denote this by $A \lhd L$ or $L \rhd A$. A is a maximal ideal of L if and only if the difference alegbra $L - A$ is *simple*. A *normal series* of L is a series of algebras starting with (0) and ending with L, each an ideal in the following. If A and B are members of a normal series of L, A contained in B, then by a *normal series from A to B* we mean a series of algebras starting with A and ending with B, each an ideal of the following. Any normal series of L can be refined by addition of terms to be a *composition series* of L, that is, a normal series in which each algebra is a maximal ideal in the following. A *composition series from A to B* is similarly defined. The difference algebras of a normal series of L are called the *factors* of the normal series; those of a composition series of L are called *composition factors* or *simple factors* of L. From the Jordan-Hölder theorem [2] we know that the factors in a composition series from A to B are determined up to isomorphism.

If A and B are subspaces of L, then for $a \in A$, $b \in B$, $[A, B]$ denotes the spaces spanned by the products $[a, b]$. It is immediate from (1) and (2) that if A and B are ideals, then $[A, B]$ is an ideal. In particular the terms of the sequences

$$L \supseteq L^2 = [L, L] \supseteq L^3 = [L^2, L] \supseteq L^4 = [L^3, L] \supseteq \cdots$$

[2] See for instance Zassenhaus [8], p. 14.

and

$$L \supseteq L' = [L, L] \supseteq L'' = [L', L'] \supseteq L^{(3)} = [L'', L''] \supseteq \cdots$$

are ideals of L. The first of these is called the *power sequence* of L; the second, the *sequence of derived algebras* of L. If the power sequence ends with (0), the algebra is *nilpotent*. If the sequence of derived algebras ends with (0), the algebra is *solvable*. $L^{(n)} \subseteq L^{2^n}$ for $n = 1, 2, \cdots$ and hence a nilpotent algebra is always solvable. The composition factors of a Lie algebra are all one-dimensional if and only if the algebra is solvable. Subalgebras and difference algebras of solvable or nilpotent Lie algebras are solvable or nilpotent respectively.

It is well known that the join of two solvable ideals is a solvable ideal; and the join of two nilpotent ideals is a nilpotent ideal. Therefore every Lie algebra has a unique maximal solvable ideal called the *radical* and a unique maximal nilpotent ideal, the *nil radical*. An algebra with zero radical is said to be *semisimple*.

A Lie algebra L is *Abelian* if $L' = [L, L] = (0)$; hence an Abelian algebra is always solvable. The *center* of a Lie algebra L is the subalgebra consisting of all elements x such that $[x, L] = (0)$. The *centralizer* of a set of elements A of L is the set $Z(A)$ of elements x of L such that $[x, A] = (0)$. It is easy to check that $Z(A)$ is always an algebra and that if A is an ideal of L then so also is $Z(A)$. It follows that the center of a Lie algebra L is always an ideal. Any nilpotent Lie algebra has a non-zero center.

If L is a Lie algebra, a linear transformation $\tilde{h}$ acting on L is a *derivation* if for $a, b \in L$,

$$[a, b]\tilde{h} = [a\tilde{h}, b] + [a, b\tilde{h}]. \tag{3}$$

It is to be noted that the derivations of L form a Lie algebra if commutation $[\tilde{h}_1, \tilde{h}_2]$ is defined to be the linear transformation $\tilde{h}_1\tilde{h}_2 - \tilde{h}_2\tilde{h}_1$. For it is easy to verify that this mapping is a derivation. The algebra of all derivations of L will be denoted by $D(L)$.

If k is an element of L, then it follows from (1) and (2) that the *adjoint mapping* which maps a into $[a, k]$ for all $a \in L$ is a derivation; this will be called an *inner derivation* and will be denoted by $\mathrm{ad}\, k$ or $\tilde{k}$. By (1) and (2)

$$[a, [k, l]] = [[a, k], l] + [k, [a, l]] = a(\tilde{k}\tilde{l} - \tilde{l}\tilde{k})$$

for $a, k, l \in L$ and therefore the inner derivations form a subalgebra of $D(L)$. The mapping k into $\tilde{k}$ is a representation of L since it is a homomorphism

of L into an algebra of linear transformations; it is called the *adjoint representation* of L. Obviously the kernel of the adjoint representation is the center of L. If L has no center then the adjoint representation is an isomorphism of L onto the algebra of inner derivations of L.

The mapping $\tilde{l}$ is *nilpotent in* L if $L\tilde{l}^m = (0)$ for some m. If $\tilde{l} = \operatorname{ad} l$ is nilpotent in L for every $l \in L$, then by Engel's theorem [3] L is nilpotent. If L is a solvable Lie algebra over a field of characteristic zero then Lie's theorem [4] states that $L' = [L, L]$ is nilpotent. Finally if L is a semisimple Lie algebra over a field of characteristic zero then by Cartan's theorem [5] $L = M_1 \oplus \cdots \oplus M_r$, where for $i = 1, \cdots, r$, M_i is a simple non-Abelian algebra which is an ideal of L.

2. Elementary properties of subinvariant subalgebras.

Definition. A is a *subinvariant subalgebra* of a Lie algebra L (or A is subinvariant in L) if there exist subalgebras A_i of L, $i = 0, 1, 2, \cdots, r$, such that $A = A_r$, $L = A_0$, and A_i is an ideal of A_{i-1}.

If A is subinvariant in L we write $A \lhd\lhd L$ or $L \rhd\rhd A$. It is obvious that a subalgebra is subinvariant if and only if it is a member of a normal series of L.

Some easy consequences of the definition of subinvariance are given in

THEOREM 1. *Let A and B be subinvariant in L, C subinvariant in A and let L^* be any subalgebra of L. Then*

(a) *C is subinvariant in L.*

(b) *$A \cap L^*$ is subinvariant in L^*.*

(c) *If $A \subseteq L^*$, then A is subinvariant in L^*.*

(d) *$A \cap B$ is subinvariant in L.*

(e) *If ν is a homomorphism of L onto $\bar{L}$, then $A\nu = \bar{A}$ is subinvariant in $\bar{L}$.*

(f) *Conversely, if $\bar{F}$ is subinvariant in $\bar{L}$, and if F is the complete inverse image of $\bar{F}$ under ν, then F is subinvariant in L.*

(g) *If L is nilpotent, then L^* is subinvariant in L.*

Proof. (a) follows directly from the definition of subinvariant subalgebra; for $C = C_s \lhd C_{s-1} \lhd \cdots \lhd C_1 \lhd A = A_r \lhd \cdots \lhd A_1 \lhd L$. To prove

[3] See Jacobson [4], p. 881.

[4] *Ibid.*, p. 880.

[5] See Cartan [1], p. 53; or Weyl [6], pp. 363 ff.

(b) we have $L = A_0 \rhd A_1 \rhd \cdots \rhd A_r = A$ and we let $(L^* \cap A_i) = A_i^*$ for $i = 0, 1, 2, \cdots, r$. Then since $A_j \rhd A_{j+1}$, it follows that $A_j^* \rhd A_{j+1}^*$, $j = 0, 1, \cdots, r-1$. For if $a_j^* \in A_j^*$, $a_{j+1}^* \in A_{j+1}^*$, then $[a_j^*, a_{j+1}^*] \in A_{j+1}^* = (L^* \cap A_{j+1})$. Therefore $L^* = A_0^* \rhd A_1^* \rhd \cdots \rhd A_r^* = (A \cap L^*)$, proving (b). Also, (c) follows immediately from (b); and (d) follows from (c) with (a).

To prove (e) we have $L = A_0 \rhd A_1 \rhd \cdots \rhd A_r = A$. If for $i = 0, 1, 2, \cdots, r$, $A_i\nu$ is denoted by $\bar{A}_i$, then $\bar{A}_i \rhd \bar{A}_{i+1}$. For if $\bar{a}_j \in \bar{A}_j$, where $j = i, i+1$, then $\bar{a}_j = a_j\nu$ for some $a_j \in A_j$ and $[\bar{a}_i, \bar{a}_{i+1}] = [a_i\nu, a_{i+1}\nu] \in \bar{A}_i$ since $[a_i, a_{i+1}] \in A_i$. It follows that $\bar{L} = \bar{A}_0 \rhd \bar{A}_1 \rhd \cdots \rhd \bar{A}_r = \bar{A}$. Conversely, to prove (f), if $\bar{L} = \bar{F}_0 \rhd \bar{F}_1 \rhd \cdots \rhd \bar{F}_r = \bar{F}$, let F_i be the complete inverse image of $\bar{F}_i$ for $i = 0, 1, \cdots, r$, and let $f_i \in F_i$. Then since $[f_i, f_{i+1}]\nu = [f_i\nu, f_{i+1}\nu] \in \bar{F}_{i+1}$, it follows that $[f_i, f_{i+1}] \in F_{i+1}$ and consequently $F_{i+1} \lhd F_i$. Then $L = F_0 \rhd F_1 \rhd \cdots \rhd F_r = F$ and $F \lhd\lhd L$, as was to be proved.

Finally, to prove (g), let L be nilpotent. Then $L^k = 0$ for some k. Consequently $L^* \lhd L^* + L^{k-1} \lhd L^* + L^{k-2} \lhd \cdots \lhd L^* + L = L$, and $L^* \lhd\lhd L$. This completes the proof of all parts of the theorem.

LEMMA 1. *If M is a maximal ideal of L and if A is subinvariant in L, then either $A \subseteq M$ or $A \cap M$ is a maximal ideal of A and $A - (A \cap M) \cong L - M$. (Here $\cong$ stands for isomorphic, $\sim$ for homomorphic onto.)*

Proof. Let $F = L - M$. Then $L \sim F$, where F is simple. Since composition series map into composition series under homomorphism, either A maps into (0), in which case $A \subseteq M$, or A maps onto F and then $(A + M) - M \cong F = L - M$. It follows from the Second Isomorphism Theorem [6] that $A - (A \cap M) \cong L - M$, proving the lemma.

As a consequence we have

THEOREM 2. *If A and B are subinvariant subalgebras of L, then the factors in a composition series from B to $A \cap B$ appear among those in a composition series from L to A.*

Proof. $L = A_0 \rhd A_1 \rhd \cdots \rhd A_r = A$, where for $i = 1, 2, \cdots, r$, A_i is a maximal ideal of A_{i-1}. Hence $B = (B \cap A_0) \geqq (B \cap A_1) \geqq \cdots \geqq (B \cap A_r) = (B \cap A)$ and the theorem is obvious.

3. A subinvariant subalgebra as the sum of an ideal and a nilpotent algebra. A key result that will be used throughout our theory of subinvariant Lie algebras is stated in

[6] See Zassenhaus [8], p. 12.

THEOREM 3. *If A is a subinvariant subalgebra of a Lie algebra L, then $A^\omega = \bigcap_{i=1}^{\infty} A^i$ is an ideal of L.*

Proof. $A^\omega = A^p$ for some integer p because of the finite dimensionality of L. It must be shown that $A^p \lhd L$. We know that $[L, A^2] \subseteq [[L, A], A]$ by (1) and (2), and we proceed by induction. If we assume for every subspace B of L that

$$[B, A^k] \subseteq [\cdots[[B, \overbrace{A], A], \cdots, A}^{k \text{ times}}]$$

then it follows that

$$[B, A^{k+1}] = [B, [A, A^k]] \subseteq [[B, A], A^k] + [[B, A^k], A]$$

$$\subseteq [\cdots [[B, A], \overbrace{A], \cdots, A}^{k \text{ factors}}] + [[\cdots [B, \overbrace{A], \cdots, A}^{k \text{ factors}}]A]$$

$$= [\cdots [B, \overbrace{A], \cdots, A}^{k+1 \text{ factors}}].$$

Hence for all n,

$$[L, A^n] \subseteq [\cdots [L, \overbrace{A], \cdots, A}^{n \text{ factors}}].$$

If n is chosen equal to $p + q$, where q is the number of factors in a composition series from L to A, then the theorem is proved.

COROLLARY 1. *If A is a subinvariant subalgebra of L such that $A = A'$, then A is an ideal of L.*

Remark. The following argument shows that the analogous statement is not true for finite groups.

Let A and B be isomorphic finite groups and consider the direct product G of A and B. Then the elements of G can be written uniquely in the form $a_i b_j$ where $a_i \in A$, $b_j \in B$. If a_j and b_j are corresponding elements under the isomorphism between A and B then the mapping $a_i b_j \to a_j b_i$ determines an automorphism of G which we will denote by σ. Obviously σ is of order 2 and if we form the subgroup H of the holomorph generated by σ and G, then G is of index 2 in H and hence is normal in H. Then it is clear that σ transforms A into B. That is, A is not normal in H. On the other hand A is normal in G and G is normal in H. Hence A is subinvariant but not invariant in H. This construction is available for an arbitrary group, in particular if the group $A = [A, A]$, where $[A, A]$ is the commutator subgroup of A, and hence we see that the analogue to Corollary 1 is not valid for finite groups.

We return to the theory of Lie algebras.

Our next task is to obtain an expression for an arbitrary subinvariant subalgebra as the sum of an ideal and a nilpotent subalgebra. To do so we first review briefly some general concepts.

If L is a vector space and $\tilde{h}$ an arbitrary linear transformation on L, we put $C = \bigcap_{i=1}^{\infty} L\tilde{h}^i$ and let $B = \bigcup_{i=1}^{\infty} B_i$, where $B_i = \{x \mid x\tilde{h}^i = 0\}$ is the linear subspace of L consisting of all elements x such that $x\tilde{h}^i = 0$. Then $C = C\tilde{h}$, $B\tilde{h}^n = 0$ for some n because of the finite dimensionality, and $L = B + C$.[7] The spaces C and B will be called the *preserved space* and *nil* space, respectively, of L relative to $\tilde{h}$, or briefly of $\tilde{h}$ if there is no danger of ambiguity. If in particular L is a Lie algebra and $\tilde{h}$ a derivation, then the nil space B of $\tilde{h}$ is an algebra. For if $b_i \in B$ for $i = 1, 2$, then $b_i\tilde{h}^{s_i} = 0$ for some s_i. It follows from Leibniz's rule that

$$[b_1, b_2]\tilde{h}^{s_1+s_2-1} = \sum_{t=0}^{s_1+s_2-1} \binom{s_1+s_2-1}{t} [b_1\tilde{h}^t, b_2 h^{s_1+s_2-1-t}] = 0$$

since either $t \geq s_1$ or $s_1 + s_2 - 1 - t \geq s_2$. Hence $[b_1, b_2] \in B$ and B is an algebra as asserted. B is called the *nil algebra* of L relative to $\tilde{h}$, or briefly, of $\tilde{h}$.

It is now possible to prove

THEOREM 4. *If L is a Lie algebra, then $L = L^\omega + H$, where H is a nilpotent algebra and $L^\omega = \bigcap_{i=1}^{\infty} L^i$.*

Proof. Let n be the dimension of L and assume the theorem true for algebras of dimension less than n. If $L\tilde{l}^n = 0$ for all $l \in L$, where $\tilde{l} = \mathrm{ad}\, l$, then L is nilpotent by Engel's theorem, and there is nothing to prove.

Otherwise there is some $l_0 \in L$ such that l_0 is not nilpotent in L. By the above remarks $L = B + C$, where C is the nil algebra of l_0 with dimension less than n and B is the preserved space. Since $B\tilde{l}_0 = B$, it follows that $B \subseteq L^\omega$. On the other hand by the induction assumption $C = C^\omega + H$, where H is a nilpotent algebra. But $C^\omega \subseteq L^\omega$. Therefore $L = B + C = B + C^\omega + H = L^\omega + H$ because $B + C^\omega \subseteq L^\omega$, and the theorem is proved.

Combining the two previous theorems we have the desired result.

THEOREM 5. *If A is a subinvariant subalgebra of a Lie algebra L, then $A = A^\omega + H$, where $A^\omega = \bigcap_{i=1}^{\infty} A^i$ is an ideal of L, and H is a nilpotent algebra.*

[7] See Jacobson [5], p. 9.

We could now pass immediately to Sections 8 and 9 and prove the tower theorem. Before doing so, however, we will first complete the general theory.

4. The radical and nil radical of a subinvariant subalgebra. This and the following three sections will deal with Lie algebras over a field of characteristic zero.

LEMMA 2. *A solvable subinvariant subalgebra of a Lie algebra over a field of characteristic zero is contained in the radical of the algebra.*

Proof. Let L be the Lie algebra with radical R and solvable subinvariant subalgebra S. If $\bar{L} = L - R$, then $\bar{L}$ is semisimple and by Cartan's theorem $\bar{L} = \bar{M}_1 \oplus \cdots \oplus \bar{M}_n$, where for $i = 1, \cdots, n$, $\bar{M}_i$ is a simple non-Abelian ideal of $\bar{L}$. It follows from the form of $\bar{L}$ that as a subinvariant subalgebra of $\bar{L}$, $\bar{S} = (S + R) - R = \bar{M}_{i_1} \oplus \cdots \oplus \bar{M}_{i_p}$, $1 \leq i_1 \leq \cdots < i_p < n$. But S is solvable and has no non-Abelian factors. Therefore $\bar{S} = 0$ and $S \subseteq R$ as the lemma asserts.

The next lemma gives results which are well known,[8] but the proof is inserted here for completeness.

LEMMA 3. *The radical R and nil radical N of a Lie algebra L over a field of characteristic zero are left invariant by every derivation $\tilde{d}$ of L; and in fact $R\tilde{d} \subseteq N$.*

Proof. Let $\tilde{d}$ be a derivation of L and form the Lie algebra $\tilde{L} = L + \tilde{d}$ where multiplication between $\tilde{d}$ and $l \in L$ is defined to be $[l, \tilde{d}] = l\tilde{d}$. From this it is obvious that $L \lhd \tilde{L}$. Let $\tilde{R}$ be the radical of $\tilde{L}$. Then $R \subseteq \tilde{R} \cap L$; for $R \subseteq L$, and by the preceding lemma $R \subseteq \tilde{R}$ since $R \lhd L \lhd \tilde{L}$. On the other hand $\tilde{R} \cap L$ is an ideal of $\tilde{L}$ since both $\tilde{R}$ and L are ideals of $\tilde{L}$. It follows that $\tilde{R} \cap L$ is an ideal of L. Since $\tilde{R}$ is solvable, $\tilde{R} \cap L$ is a solvable ideal of L and $\tilde{R} \cap L \subseteq R$. Hence $R = \tilde{R} \cap L$. As we have seen, this implies that R is an ideal of $\tilde{L}$ and therefore $R\tilde{d} \subseteq R$; and thus the radical is invariant under every derivation $\tilde{d}$ of L, as the first part of the lemma asserts.

To show that the nil radical N is invariant with respect to the derivation $\tilde{d}$ we consider the subalgebra $R_1 = R + \tilde{d}$ of $L + \tilde{d} = \tilde{L}$. This is a subalgebra since it has already been shown that $R\tilde{d} \subseteq R$. It follows that R_1 is solvable and consequently by Lie's theorem $R_1' = [R_1, R_1]$ is nilpotent.

[8] The proof given here was presented by Professor N. Jacobson to his class in non-associative algebras at Yale University. See also Hochschild [3], p. 62.

But $R_1' \lhd R_1$; hence $R_1' \lhd R$ since $R_1' \subseteq R$. Then as a nilpotent ideal of R, $R_1' \subseteq N_R$, the nil radical of R. Hence $Rd \subseteq R_1' \subseteq N_R$. By taking $\tilde{d}$ to be an inner derivation determined by the element l of L we have $[R, l] \subseteq N_R$. Since this is true for all $l \in L$, $[R, L] \subseteq N_R$. Hence $[N_R, L] \subseteq N_R$ and N_R is a nilpotent ideal of L. Then by the definition of nil radical $N_R \subseteq N$. Obviously $N_R \supseteq N$ and thus $N = N_R$. But we had $R\tilde{d} \subseteq N_R$. Hence $R\tilde{d} \subseteq N$ and consequently also $N\tilde{d} \subseteq N$, completing the proof of the lemma.

LEMMA 4. *If M is a nilpotent subinvariant subalgebra of L, a Lie algebra over a field of characteristic zero, then M is contained in N, the nil radical of L.*

Proof. $M \lhd \lhd L$ and therefore $M \lhd M_1 \lhd M_2 \lhd \cdots \lhd M_n = L$. For $j = 1, \cdots, n$, let N_j be the nil radical of M_j and let $m_j \in M_j$. Then $M \subseteq N_1$ since M is a nilpotent ideal of M_1. But for $i = 1, \cdots, n-1$, $[M_i, m_{i+1}] \subseteq M_i$ and therefore m_{i+1} defines a derivation $\tilde{m}_{i+1}$ in M_i. It follows from the preceding lemma that $N_i\tilde{m}_{i+1} \subseteq N_i$, whence $N_i \lhd M_{i+1}$ and consequently $N_i \subseteq N_{i+1}$. Then $M \subseteq N_1 \subseteq N_2 \subseteq \cdots \subseteq N_n = N$, proving the lemma.

We can now prove the main result of this section.

THEOREM 6. *Let L be a Lie algebra over a field of characteristic zero, with radical R and nil radical N. Let A be a subinvariant subalgebra of L with radical R_A and nil radical N_A. Then $R_A = A \cap R$ and $N_A = A \cap N$.*

Proof. $A \cap R$ is a solvable ideal of A since $R \lhd L$ and hence $A \cap R \subseteq R_A$. On the other hand $R \subseteq A$ and, by Lemma 2, $R_A \subseteq R$. Hence $R_A \subseteq A \cap R$ and $R_A = A \cap R$. In like manner Lemma 4 gives that $N_A = A \cap N$ and the proof is complete.

5. The sum theorems.

THEOREM 7. *Let A and B be subinvariant subalgebras of L, a Lie algebra over a field of characteristic zero. If $\{A, B\}$ denotes the algebra generated by A and B, then $\{A, B\}$ is subinvariant in L.*

Proof. If both A and B are nilpotent then, by Lemma 4, A, B and consequently $\{A, B\}$ are contained in the nil radical N of L. Then by Theorem 1 (g), $\{A, B\}$ is subinvariant in N and therefore in L. Thus in this case the theorem is true.

In a general case we have $(0) \neq A^\omega$ or $(0) \neq B^\omega$. Therefore, by

Theorem 3, $\{A^\omega, B^\omega\}$ is a non zero ideal of L. Let $\bar{A} = (A + \{A^\omega + B^\omega\}) - \{A^\omega, B^\omega\}$, $\bar{B} = (B + \{A^\omega, B^\omega\}) - \{A^\omega, B^\omega\}$. Then $\bar{A}$ and $\bar{B}$ are nilpotent in $\bar{L} = L - \{A^\omega, B^\omega\}$ and by the first part of the proof $\{\bar{A}, \bar{B}\} \lhd\lhd \bar{L}$. But $\{\bar{A}, \bar{B}\} = \overline{\{A, B\}}$ where $\overline{\{A, B\}} = \{A, B\} - \{A^\omega, B^\omega\}$. For if C_i stands for A or B, then $\overline{\{A, B\}} = \overline{\Sigma[\cdots[C_1, C_2], \cdots, C_n]} = \Sigma[\cdots[\bar{C}_1, \bar{C}_2], \cdots, \bar{C}_n] = \{\bar{A}, \bar{B}\}$. Hence $\{\bar{A}, \bar{B}\} = \overline{\{A, B\}}$ and by the above $\overline{\{A, B\}} \lhd\lhd \bar{L}$. It follows from Theorem 1 (f) that $\{A, B\} \lhd\lhd L$ and the proof is complete.

In like manner we prove the following result.

THEOREM 8. *Let A be subinvariant in L, a Lie algebra over a field of characteristic zero. Let b be any element of L and let A^* be the smallest algebra containing A such that $[A^*, b] \subseteq A^*$. Then A^* is subinvariant in L.*

Proof. We first note that $A^* = \{\sum_{k=0}^{\infty} A\tilde{b}^k\}$. For certainly $\{\sum_{k=0}^{\infty} A\tilde{b}^k\}$ contains A and is invariant relative to $\tilde{b}$ since the space spanned by $\sum_{k=0}^{\infty} A\tilde{b}^k$ is left invariant by $\tilde{b}$. Hence $\{\sum_{k=}^{\infty} A\tilde{b}^k\} \supseteq A^*$. On the other hand $A^* \supseteq A$ and since $[A^*, b] \subseteq A^*$, $A^* \supseteq A\tilde{b}^k$ for all k. It follows that $A^* \supseteq \{\sum_{k=0}^{\infty} A\tilde{b}^k\}$. Consequently $A^* = \{\sum_{k=0}^{\infty} A\tilde{b}^k\}$.

If now A is nilpotent then $A \subseteq N$, the nil radical of L. But N is an ideal of L and therefore by its form $A^* \subseteq N$ and consequently A^* is subinvariant in L. If A is not nilpotent then A^ω is a non zero ideal of L. Let $\bar{L} = L - A^\omega$, $\bar{A} = A - A^\omega$, $\overline{A^*} = A^* - A^\omega$, and $\bar{b} = b + A^\omega$. Then $\overline{A^*} = \{\sum_{k=0}^{\infty} A\tilde{b}^k\} = \{\sum_{k=0}^{\infty} \bar{A}\tilde{\bar{b}}^k\}$. But $\bar{A}$ is in the nil radical of $\bar{L}$ and hence so also is $\{\sum_{k=0}^{\infty} \bar{A}\tilde{\bar{b}}^k\} = A^*$. It follows that $\overline{A^*} \lhd\lhd \bar{L}$ and hence $A^* \lhd\lhd L$, proving the theorem.

From Theorem 2 we know that if A and B are subinvariant then the factors in a composition series between A and $(A \cap B)$ occur among those in a composition series between $\{A, B\}$ and B. The converse of this, namely that the factors in a composition series from $\{A, B\}$ to B occur among those in a composition series from A to $(A \cap B)$ is true if we add the qualifying statement "not counting multiplicities of one-dimensional factors," and if we assume that the base field has characteristic zero. To show this we need the following result.

LEMMA 5. *If L is a Lie algebra over a field of characteristic zero, then $L = L^{(\omega)} + R$, where $L^{(\omega)} = \bigcap_{i=1}^{\infty} L^{(i)}$ and R is the radical of L.*

Proof. $L - R$ is semisimple. Therefore $L - R = [L - R, L - R]$

$= \cdot \cdot \cdot = (L - R)^{(\omega)}$ and $L - R = (L^{(\omega)} + R) - R$. Consequently $L = L^{(\omega)} + R$ as was to be shown.

It is interesting to contrast this lemma with Theorem 5. If A is subinvariant in L, and if R_A is the radical of A, then the lemma tells us that A is the sum of two algebras which are subinvariant in L. On the other hand if the properties of A alone are being considered, then the expression $A^{\omega} + H$ given by Theorem 5 is useful because of the nilpotency of H.

We will now prove the partial converse of Theorem 2.

THEOREM 9. *If A and B are subinvariant in L, a Lie algebra over a field of characteristic zero, then the factors in a composition series from $\{A, B\}$ to A occur among those in a composition series from B to $(A \cap B)$, not counting multiplicities of one-dimensional factors.*

Proof. We observe first that the composition factors are never one-dimensional or Abelian between a Lie algebra and its radical and are always one-dimensional for a solvable algebra. Now by the corollary to Theorem 3, $A^{(\omega)} = \bigcap_{i=1}^{\infty} A^{(i)}$ is an ideal of L. Let $\bar{L} = L - A^{(\omega)}$, $\bar{A} = A - A^{(\omega)}$, $\bar{B} = (B + A^{(\omega)}) - A^{(\omega)}$, $\overline{\{A, B\}} = \{A, B\} - A^{(\omega)}$, and $\overline{(A \cap B)} = ((A \cap B) + A^{(\omega)}) - A^{(\omega)}$. We note that, as in the proof of Theorem 7, $\overline{\{A, B\}} = \{\bar{A}, \bar{B}\}$ and $\overline{(A \cap B)} = \bar{A} \cap \bar{B}$. Then it is well known that the composition factors from $\{A, B\}$ to A are the same as those from $\overline{\{A, B\}}$ to $\bar{A}$. Also the factors between $\bar{B}$ and $\overline{(A \cap B)}$ occur among those between B and $A \cap B$. Hence it is sufficient to prove the theorem for $\bar{A}$, $\bar{B}$ in $\bar{L}$.

In this case $\bar{A}$ and consequently $(\bar{A} \cap \bar{B})$ are solvable. By the preceding lemma $\bar{B} = \bar{B}^{(\omega)} + \bar{R}_B$ where $\bar{R}_B$ is the radical of $\bar{B}$ and $\bar{B}^{(\omega)} = \bigcap_{i=1}^{\infty} \bar{B}^{(i)}$. Then $\bar{R}_B$ and $\bar{A}$ are contained in $\bar{R}$, the radical of $\bar{L}$. Hence $\overline{\{A, B\}} = \{\bar{A}, \bar{B}\}$ $= \{\bar{A}, \bar{B}^{(\omega)}, \bar{R}_B\} \subseteq \{\bar{B}^{(\omega)}, \bar{R}\} = \bar{B}^{(\omega)} + \bar{R}$ since each of the latter algebras is an ideal of $\bar{L}$. Now the non-Abelian composition factors of $\bar{B}^{(\omega)} + \bar{R}$ are the same as those of $(\bar{B}^{(\omega)} + \bar{R}) - \bar{R}$, and consequently are contained in $\bar{B}^{(\omega)}$ since $(\bar{B}^{(\omega)} + \bar{R}) - \bar{R}$ is a homomorphic image of $B^{(\omega)}$. It follows that the non-Abelian composition factors of $\{\bar{A}, \bar{B}\}$ must occur among those of $\bar{B}^{(\omega)}$, consequently in $\bar{B}$; and hence between $\bar{B}$ and $\overline{(A \cap B)}$, since $\overline{(A \cap B)}$ is solvable, and therefore in $\bar{R}_B$.

There are two possible cases in considering the one-dimensional factors. First there may be one or more one-dimensional factors between $\bar{B}$ and $(\bar{A} \cap \bar{B})$; in this case the theorem is obviously true. On the other hand, if none of the factors from $\bar{B}$ to $(\bar{A} \cap \bar{B})$ is one-dimensional then $\bar{R}_B = (\bar{A} \cap \bar{B})$,

since the latter is solvable and subinvariant and hence contained in $\bar{R}_B$. In this case, $\{\bar{A}, \bar{B}\} = \{\bar{A}, \bar{B}^{(\omega)}, \bar{R}_B\} = \{\bar{A}, \bar{B}^{(\omega)}\} = \bar{A} + \bar{B}^{(\omega)}$, since $\bar{B}^{(\omega)}$ is an ideal of $\bar{L}$. Then $\bar{A}$ is the radical of $\bar{A} + \bar{B}^{(\omega)}$. For certainly $\bar{A}$ is contained in $\bar{S}$, the radical of $\bar{A} + \bar{B}^{(\omega)}$. On the other hand, if $\bar{s} \in \bar{S}$, then $\bar{s} = \bar{a} + \bar{b}$, where $\bar{a} \in \bar{A}$, $\bar{b} \in \bar{B}^{(\omega)}$. But then since $\bar{a} \in \bar{S}$, $\bar{b} \in \bar{S}$ and $\bar{b} \in \bar{R}_B \subseteq \bar{A}$; hence $\bar{S} \subseteq \bar{A}$. It follows that $\bar{A}$ is the radical of $\bar{A} + \bar{B}^{(\omega)} = \{\bar{A}, \bar{B}\}$ and the factors from $\{\bar{A}, \bar{B}\}$ to $\bar{A}$ are non-Abelian. This completes the proof of the theorem.

6. The minimal subinvariant subalgebra determined by a single element. It is clear that there is a unique minimal subinvariant subalgebra determined by an arbitrary element of a Lie algebra because of the finite dimensionality and because the intersection of two subinvariant algebras is subinvariant.

We shall consider first the minimal subinvariant subalgebra determined by an element of the radical. For this we need the following result.

LEMMA 6. *Let a be an element of the Lie algebra L over an arbitrary field, and let L_1 be the preserved space of $\tilde{a}$. If $a \in A$, A subinvariant in L, then L_1 is contained in A.*

Proof. $L_1 = L_1\tilde{a}^k$, which for k large enough is in A, since $a \in A$, and A is subinvariant in L.

We can now prove the following theorem.

THEOREM 10. *Let R be the radical of L, a Lie algebra over a field of characteristic zero, and let L_1 be the preserved space in L of the derivation $\tilde{a}$ determined by an element a in R. Let L_a be the algebra generated by L_1 and let $L_a + (a) = A$. Then A is the minimal subinvariant subalgebra of L containing a, and L_a is an ideal of L. L_a will be called the* associated ideal *to a.*

Proof. In order to prove that A is an algebra it is sufficient to observe that if $\tilde{a} = \operatorname{ad} a$, then $L_1\tilde{a} = L_1$ and consequently $L_a\tilde{a} \subseteq L_a$. We observe next that $L_1 = L_1\tilde{a} \subseteq R$, since $a \in R$ and R is an ideal. It follows that $L_1 = L_1\tilde{a} \subseteq R' \subseteq N$, the nil radical of L, and consequently $L_a \subseteq N$. Then for all s, $[N^s, L_a] \subseteq N^{s+1}$. Let $N^s = N_{0,s} \oplus N_{1,s}$ where $N_{0,s}$ is the nil space, $N_{1,s}$ the preserved space of N^s relative to $\tilde{a}$. Then $N_{1,s} \subseteq L_1 \subseteq L_a$ and $N_{0,s}\tilde{a}^{k_s} = 0$ for k_s depending on s. If $N^t = 0$ we have

$$\begin{aligned} L_a + (a) = A &\lhd A + N_{0,t-1}\tilde{a}^{k_{t-1}-1} \lhd A + N_{0,t-1}\tilde{a}^{k_{t-1}-2} \\ &\lhd \cdots \lhd A + N_{0,t-1} = A + N^{t-1} \lhd A + N^{t-1} + N_{0,t-2}\tilde{a}^{k_{t-2}-1} \\ &\lhd \cdots \lhd A + N \lhd R \lhd L \end{aligned}$$

where each step is true because of the facts that $A = L_a + (a)$, $[N^s, L_a] \subseteq N^{s+1}$, and $[N_{0,t-r}\tilde{a}^{k_{t-r}-p_r}, a] = N_{0,t-r}\tilde{a}^{k_{t-r}-p_{r+1}}$ for $0 < r < t$ and $0 < p_r < k_{t-r}$. This proves that A is subinvariant in L. A is the minimal subinvariant subalgebra containing a by the previous lemma.

To show that L_a is an ideal of L we show that $L_a = \bigcap_{i=1}^{\infty} A^i$, as follows. $L_a \supseteq \bigcap_{i=1}^{\infty} A^i$; for $L_a \supseteq A^2$, since $A = L_a + (a)$ and $L_a\tilde{a} \subseteq L_a$. On the other hand, $L_1 = L_1\tilde{a} \subseteq \bigcap_{i=1}^{\infty} A^i$ and consequently $L_a \subseteq \bigcap_{i=1}^{\infty} A^i$, since L_a is the algebra generated by L_1. Then by Theorem 3, L_a is an ideal of L, and the proof of the theorem is complete.

We observe that if $a \in N$, then $L_1 = 0$, $A = (a)$ and the theorem is true by virtue of Theorem 1 (g).

The next task is to study the subinvariant subalgebras which map onto the simple components of a Lie algebra modulo its radical. The result that we shall need is contained in the following theorem.

THEOREM 11. *Let L be a Lie algebra over a field of characteristic zero with radical R; and let $\bar{L} = L - R = \sum_{i=1}^{n} \oplus \bar{M}_i$, where for $i = 1, \cdots, n$, $\bar{M}_i$ is a simple algebra and an ideal of $\bar{L}$. If ν is the natural mapping of L onto $\bar{L}$, let T_i be the complete inverse image under ν of $\bar{M}_i$. Then $T_i^{(\omega)} = \bigcap_{k=1}^{\infty} T_i^{(k)}$ is an ideal of L such that $T_i^{(\omega)}\nu = \bar{M}_i$, and $T_i^{(\omega)}$ is contained in every subvariant subalgebra of L which maps under ν onto $\bar{M}_i$.*

Proof. Since $\bar{M}_i$ is an ideal of $\bar{L}$, it follows that T_i and consequently $T_i^{(\omega)}$ are ideals of L. Now, T_i contains R and R is the radical of T_i. By its definition $T_i - R = \bar{M}_i$, and therefore in a composition series of T_i there is only one simple non-Abelian factor, and this factor is isomorphic to $\bar{M}_i$. It follows that if A_i is any subinvariant subalgebra of T_i, then a simple non-Abelian factor must occur either in a composition series from T_i to A_i or in that of A_i, but not in both series. In the former case A_i is solvable and hence is contained in R; in the latter case A_i is not solvable and $A_i\nu = \bar{M}_i$.

Suppose now that A_i and B_i are subinvariant subalgebras of L which map under ν onto $\bar{M}_i$. Then A_i and B_i are contained in T_i, since T_i is the complete inverse image of $\bar{M}_i$; and the simple factors between T_i and A_i and between T_i and B_i are all one-dimensional, since $A_i\nu = B_i\nu = \bar{M}_i$. By Theorem 2 it follows that the simple factors between A_i and $(A_i \cap B_i)$ are one-dimensional and hence the same result holds for the simple factors between T_i and $(A_i \cap B_i)$. It follows that $(A_i \cap B_i)\nu = \bar{M}_i$. From this

together with the facts that $(A_i \cap B_i)$ is subinvariant in L and that L is finite-dimensional we know that there exists a subinvariant subalgebra V_i of L such that $V_i\nu = \bar{M}_i$ and such that $V_i \subseteq A_i$ if A_i is any subinvariant subalgebra of L whose image under ν is $\bar{M}_i$. Now, $T_i^{(\omega)}\nu = \bar{M}_i$ since the simple factors between T_i and $T_i^{(\omega)}$ are one-dimensional and $T_i^{(\omega)}$ is subinvariant, in fact, an ideal of L. Hence $T_i^{(\omega)} \supseteq V_i$. But then the factors between $T_i^{(\omega)}$ and V_i are one-dimensional since $T_i^{(\omega)}\nu = V_i\nu = \bar{M}_i$. This implies that $T_i^{(\omega)} = V_i$ and the theorem is proved.

We note that $(\sum_{i=1}^{n} T_i^{(\omega)})\nu = \sum_{i=1}^{n} \bar{M}_i = \bar{L}$ and hence the composition factors between L and $\sum_{i=1}^{n} T_i^{(\omega)}$ are all one-dimensional. It follows that $L^{(\omega)} \subseteq \sum_{i=1} T_i^{(\omega)}$. On the other hand $(\sum_{i=1} T_i^{(\omega)})' = \sum_{i=1} T_i^{(\omega)}$ and hence $L^{(\omega)} = \sum_{i=1}^{n} T_i^{(\omega)}$. Then from Lemma 5 we also have that $L = \sum_{i=1}^{n} T_i^{(\omega)} + R$.

If now l is arbitrary in L, let $l\nu = \sum_{j\in J} \bar{m}_j$, where $J = (i_1, \cdots, i_r)$ and where $\bar{m}_j \in \bar{M}_j$, $\bar{m}_j \neq 0$. It is possible to choose $b \in \sum_{j\in J} T_j^{(\omega)}$ such that $b\nu = l\nu$, since $\sum_{j\in J} T_j^{(\omega)}$ maps onto $\sum_{j\in J} \bar{M}_j$. Hence $l - b = a \in R$ and we let L_a be the associated ideal of a. The characterization of the minimal subinvariant subalgebra determined by l is then given in the following theorem.

THEOREM 12. *Let L be a Lie algebra satisfying the hypotheses of the preceding theorem. Let A be the minimal subinvariant subalgebra determined by an arbitrary element l of L. Then in the notation above,*

$$A = \sum_{j\in J} T_j^{(\omega)} + L_a + (a).$$

Proof. $A\nu$ contains $l\nu = \sum_{j\in J} \bar{m}_j$, where $\bar{m}_j \neq 0$. Hence $A\nu$ contains $\sum_{j\in J} \bar{M}_j$, since the only ideals and therefore the only subinvariant subalgebras of a semisimple algebra are sums of the simple components. It follows from Theorem 11 that A contains $\sum_{j\in J} T_j^{(\omega)}$. Then $b \in A$ and consequently $l - b = a \in A$. Then by Theorem 10, $L_a \subseteq A$. Therefore $A \supseteq \sum_{j\in J} T_j^{(\omega)} + L_a + (a)$. But the latter expression represents a subinvariant algebra containing l. For $L_a + (a)$ is subinvariant by Theorem 10; and $\sum_{j\in J} T_j^{(\omega)}$ is an ideal. Hence by the minimality of A, $A = \sum_{j\in J} T_j^{(\omega)} + L_a + (a)$, and the theorem is proved.

7. The greatest subinvariant subalgebra contained in, and the least subinvariant subalgebra containing, an arbitrary algebra. If U is an arbitrary subalgebra of a Lie algebra L, let U_* denote the algebra generated by all the subinvariant subalgebras of L contained in U; and let U^* denote the

intersection of all the subinvariant subalgebras of L containing U. Then by Theorem 1 (d), U^* is minimal such that $U^* \lhd \lhd L$ and $U^* \supseteq U$. If the base field of the algebra is of characteristic zero, then by Theorem 7, U_* is maximal such that $U_* \lhd \lhd L$ and $U_* \subseteq U$. We observe that $U_* \subseteq U \subseteq U^*$.

The following results analogous to those for finite groups are then valid.

LEMMA 7. *If U is any subalgebra of a Lie algebra L over a field of characteristic zero, then U_* is an ideal of U.*

Proof. Let b be any element of U. If U_1 is the minimal algebra such that $U_1 \supseteq U_*$ and $[U_1, b] \subseteq U_1$ then it follows from Theorem 8 that U_1 is subinvariant in L. Obviously $U_1 \subseteq U$ and therefore $U_1 \subseteq U_*$. It follows that $[U_*, b] \subseteq U_*$ for all $b \in U$ and therefore U_* is an ideal of U.

THEOREM 13. *If U and V are subalgebras of a Lie algebra L over an arbitrary field and if $\{U, V\} = U + V$, then $\{U^*, V^*\} = U^* + V^*$.*

Proof. It will be sufficient to show that $U^* = U + U^{*\omega}$ and $V^* = V + V^{*\omega}$, since $U^{*\omega}$ and $V^{*\omega}$ are ideals of L. Obviously $U^* \supseteq U + U^{*\omega}$. On the other hand, $(U + U^{*\omega}) - U^{*\omega}$ is a subalgebra of the nilpotent algebra $U^* - U^{*\omega}$. Hence by Theorem 1 (g), $(U + U^{*\omega}) - U^{*\omega}$ is subinvariant in $U^* - U^{*\omega}$, which by Theorem 1 (e) is subinvariant in $L - U^{*\omega}$. Then by Theorem 1 (f), $U + U^{*\omega}$ is subinvariant in L and therefore $U + U^{*\omega} \supseteq U^*$. It follows that $U^* = U + U^{*\omega}$. Likewise $V^* = V + V^{*\omega}$ and the proof is complete.

The following example shows that $\{U, V\} = U + V$ need not imply that $\{U_*, V_*\} = U_* + V_*$. Let $L = S \oplus N$ where S is semisimple and N is the nilpotent algebra with basis x, y, z such that $[x, y] = z$, $[x, z] = [y, z] = 0$; and let N and S be ideals of L. Then if s is an element of S, consider the subalgebras $U = \{s, x\}$, $V = \{y, s - z\}$. Then $\{U, V\} = \{s, x, y, s - z\} = (x, y, z, s) = U + V$. But $U_* = (x)$, $V_* = (y)$; and $\{U_*, V_*\} = \{x, y\} = (x, y, z) \neq U_* + V_*$.

We now give a generalization of Theorem 7 to the case where one of the algebras need not be subinvariant.

THEOREM 14. *Let A be subinvariant in a Lie algebra over a field of characteristic zero. Let V be any subalgebra of L, and let $K = \{A, V\}$. Then the composition factors from K^* to K_* are a part of those from V^* to V_*.*

Proof. By Lemma 7, K_* is an ideal of K. Therefore since $V \subseteq K$, $\{K_*, V\} = K_* + V$. Then by the preceding theorem $\{K_*, V^*\} = K_* + V^*$.

It follows that the composition factors from $K_* + V^*$ to K_* are the same as those from V^* to $V^* \cap K_*$. For by Theorem 2 the composition factors from V^* to $V^* \cap K_*$ occur those from $K_* + V^*$ to K_*. On the other hand if two algebras of a series from $K_* + V^*$ to K_* have the same intersections with V^*, then by their forms they must be the same algebras and hence there is a one-to-one correspondence between the composition factors of the two series, $K_* + V^*$ to K_*, and V^* to $V^* \cap K_*$.

Now $V \subseteq K$ and therefore $V_* \subseteq V^* \cap K_*$. Also $A \subseteq K_*$, $V \subseteq V^*$; consequently $K = \{A, V\} \subseteq \{K_*, V^*\} = K_* + V^*$ and $K^* \subseteq K_* + V^*$. We then have that $V_* \subseteq V^* \cap K_* \subseteq V^*$ and $K_* \subseteq K^* \subseteq K_* + V^*$ and it is easy to check that the theorem follows.

Now let k be any element of a Lie algebra L and denote the nil algebra determined by the derivation $\tilde{k}$ by K. We recall that k is called a *regular element* if $\dim K$ is minimal for all the elements k of L; and in this case K is called a *Cartan subalgebra*. The theory of these subalgebras is fundamental in the general structure theory of Lie algebras. The following result may therefore be of interest.

THEOREM 15. *If K is a Cartan subalgebra of a Lie algebra L over an arbitrary field, then* (1) $K^* = L$. (2) K_* *need not be an ideal of* L.[9]

Proof. Let K be the nil algebra of k. Then $L = K \oplus L_1$, where L_1 is the preserved space of $\tilde{k}$. By Lemma 6, $K^* \supseteq L_1$. Hence $K^* \supseteq L_1 + K = L$ and $K^* = L$, proving (1). To prove (2) we construct the algebra $A = \{a_1, a_2, b, k\}$ for which a partial multiplication table is defined to be $[a_1, a_2] = 0$, $[a_1, b] = 0$, $[a_1, k] = a_1$, $[a_2, b] = a_1$, $[a_2, k] = a_2$, $[b, k] = 0$; the other products are defined in the obvious way and it is easy to see that A is an algebra. The subalgebra $K = \{b, k\} = (b, k)$ is the nil algebra of $\tilde{k}$ and is a Cartan subalgebra. For otherwise there would be an element whose adjoint mapping determines a one-dimensional nil algebra and hence a preserved space of dimension three. But this is impossible since $A' = [A, A] = (a_1, a_2)$ is of dimension two. Therefore K is a Cartan subalgebra as asserted. Now K_* cannot contain k, else by Lemma 6 K_* would contain the preserved space of $\tilde{k}$, which is (a_1, a_2). On the other hand $B = (b) \lhd \{b, a_1\} \lhd \{b, a_1, a_2\} \lhd A$, and therefore $B = K_*$. But B is not an ideal of A, for $[b, a_2] = -a_1 \notin B$; and the proof is complete.

The above algebra A also illustrates that the centralizer of a subinvariant

[9] Compare Wielandt's result [7], p. 229. If S is a Sylow subgroup of a finite group G, then S^* and S_* are normal subgroups of G.

need not be subinvariant. For if Z is the centralizer of B, then $Z = (a_1, b, k)$. But $[a_2, k] = a_2$ and therefore Z is not subinvariant in A.

8. The centralizer of a subinvariant subalgebra. The object of this and the following sections will be to prove the tower theorem for Lie algebras. The results depend only on those of Section 1 through 3 and are valid for an arbitrary base field.

We begin with two lemmas.

LEMMA 8. *If H is a nilpotent Lie algebra with center Q, and if $P \neq (0)$ is an ideal of H, then $(P \cap Q) \neq (0)$.*

Proof. For some n,

$$[\cdots[P, \overbrace{H], \cdots, H}^{n-1 \text{ times}}] \neq (0), [\cdots[P, \overbrace{H], \cdots, H}^{n \text{ times}}] = (0),$$

since H is nilpotent. Therefore $[\cdots[P, \overbrace{H], \cdots, H}^{n-1 \text{ times}}] \subseteq Q$. But $[\cdots[P, \overbrace{H], \cdots, H}^{n-1 \text{ times}}] \subseteq P$ since P is an ideal. Therefore $P \cap Q \neq (0)$ as was asserted.

LEMMA 9. *Let A be a Lie algebra over an arbitrary field and let $A^\omega = \bigcap_{k=1}^{\infty} A^k$. If Z is the centralizer of A^ω in A, and if Z is not contained in A^ω, then A has a center not equal to zero.*

Proof. A^ω is an ideal of A and consequently so also is Z. By Theorem 4, $A = A^\omega + H$, where H is a nilpotent algebra. Let $A_1 = Z + H$. Then A_1 is an algebra, since Z is an ideal and H is an algebra. Hence $A_1 = A_1^\omega + H_1$, where H_1 is a nilpotent algebra. But $A_1^\omega = (Z + H)^\omega \subseteq Z$ by the nilpotency of H and because Z is an ideal. If $Z \subseteq A_1^\omega$, then $Z \subseteq A^\omega$, since $A_1^\omega \subseteq A^\omega$. But this contradicts the hypothesis that $Z \not\subseteq A^\omega$. It follows that there is a $z_1 \in Z$, $z_1 \notin A_1^\omega$. We shall write $z_1 = a_1^\omega + h_1$, where $a_1 \in A_1^\omega$ and $h_1 \in H_1$. This is permissible since $Z \subseteq A_1^\omega + H_1$. Now $h_1 \neq 0$, else $z_1 \in A_1^\omega$, contrary to the choice of z_1. Hence $h_1 = z_1 - a_1^\omega \in Z$, since $z_1 \in Z$ and $a_1^\omega \in A_1^\omega \subseteq Z$. It follows that $H_1 \cap Z \neq (0)$.

Now $H_1 \cap Z$ is an ideal of H_1, since Z is an ideal of A. H_1 has a center $Q \neq 0$, since H_1 is nilpotent. Consequently by the previous lemma $P = (H_1 \cap Z) \cap Q \neq 0$. Finally, P is in the center of A. For $A = A^\omega + H$,

where $H \subseteq A_1^\omega + H_1$; and hence $H \subseteq A^\omega + H_1$, since $A_1^\omega \subseteq A^\omega$. Consequently $A = A^\omega + H_1$. But P is in the centralizer of A^ω since $P \subseteq Z$, and P is in the center of H_1 since $P \subseteq Q$. Therefore P is in the center of A, proving the lemma.

We can now prove the following theorem.

THEOREM 16. *Let L be a Lie algebra over an arbitrary field. If A is subinvariant in L and if A has centralizer (0), then the centralizer Z of $A^\omega = \bigcap_{k=1}^{\infty} A^k$ is contained in A^ω.*

Proof. $A \cap Z$ is contained in A^ω for if not, then by the previous lemma, A would have a center not equal to zero, contrary to hypothesis. Suppose Z is not contained in A^ω; then by the above remark Z is not contained in A. Now A^ω is an ideal of L and consequently so also is Z. It follows that $A + Z = K$ is an algebra, and $A \neq K$ since $Z \not\subseteq A$. Then since $A \lhd\lhd L$, $A \lhd A_1 \lhd\lhd K$, where $A_1 \neq A$. Let a_1 be an element of A_1, $a_1 \notin A$. By the form of K, $a_1 = a + z_1$, where $a \in A$, $z_1 \in Z$, and $z_1 \notin A$ since $a_1 \notin A$. Let $B = A + (z_1)$. B is an algebra since A is an ideal of A_1. Let $B^\omega = \bigcap_{k=1}^{\infty} B^k$. Then $B^\omega = A^\omega$. For $B^k \supseteq A^k$ for all k since $B \supset A$. On the other hand, $[A, z_1] \subseteq A \cap Z$ since A is an ideal of A_1 and Z is an ideal of L. But $A \cap Z \subseteq A^\omega$, and hence $B^k \subseteq A^k$ for all $k > 1$. Then since $z_1 \notin A^\omega$, by the previous lemma we know that B has a center and consequently A has a nonzero centralizer in L. But this is contrary to hypothesis and hence Z must be contained in A^ω, as the theorem asserts.

9. The Tower theorem. Before we state the tower theorem it is convenient to have the following lemma.

LEMMA 10. *Let L be a Lie algebra with center zero, and let D_1 be the derivation algebra of L, D_0 the algebra of all inner derivations of L. Then* (1) *L is isomorphic to D_0;* (2) *D_0 is an ideal of D_1;* (3) *the centralizer of D_0 in D_1 is zero.*

Proof. Since L has center zero, it is evident that the representation of L by D_0 is an isomorphism, proving (1). To show that D_0 is an ideal of D_1, we let $\tilde{d} \in D_0$ correspond to $d \in L$; that is, for $x \in L$, $x\tilde{d} = [x, d]$. Then for arbitrary $\tilde{d}_1 \in D_1$, $x[\tilde{d}_1, \tilde{d}] = x(\tilde{d}_1\tilde{d} - \tilde{d}\tilde{d}_1) = [x\tilde{d}_1, d] - [x, d]\tilde{d}_1$ $= [x\tilde{d}_1, d] - [x\tilde{d}_1, d] - [x, d\tilde{d}_1] = -[x, d\tilde{d}_1]$. Hence $[\tilde{d}_1, \tilde{d}]$ is an inner derivation, which means that D_0 is an ideal of D_1, proving (2).

Finally, if $\tilde{d}_1$ is in the centralizer of D_0 in D_1, then $[\tilde{d}_1, \tilde{d}] = 0$ for all $\tilde{d} \in D_0$; whence by the above calculation $[x, d\tilde{d}_1] = 0$ for all x. But L has center zero. Therefore $d\tilde{d}_1 = 0$ for all $d \in L$. Hence $\tilde{d}_1 = 0$, proving the last statement of the lemma.

It follows from the lemma that if L has center zero, then $L \cong D_0$, $D(L)$ has no center, and $L \lhd D(L)$. Accordingly we can make the following definition.

Definition. If L is a Lie algebra with center zero, let $D_0 = L$, $D_1 = D(L)$, the derivation algebra of $L, \cdots, D_n = D(D_{n-1}), \cdots$, then the series of algebras $D_0, D_1, D_2, \cdots$ is called the *tower of derivation algebras* of L.

We wish to find a bound for the dimensions of the algebras D_k, $k = 1, 2, \cdots$, of the tower of derivation algebras of an arbitrary Lie algebra with center zero. The result is given in the following theorem.

THEOREM 17. *Let L be a Lie algebra over an arbitrary field and let the center of L be zero, so that $L = D_0$. Then $D_0, D_1, \cdots, D_n$, the tower of derivation algebras of L, ends finitely, and* $\dim D_n \leq d + c$, *where d is the dimension of the derivation algebra of $L^\omega = \bigcap_{k=1}^{\infty} L^k$, and c is the dimension of the center of L^ω.*

Two lemmas are needed in the proof.

LEMMA 11. *If A is an ideal of L and Z is the centralizer of A in L, then* $\dim L \leq \dim Z + d$, *where* $d = \dim D(A)$, *the derivation algebra of A.*

Proof. For $l \in L$, $[A, l] \subseteq A$, since A is an ideal of L. It follows that $\tilde{l} = \operatorname{ad} l$ acting in A is a derivation of A. Consequently there are at most d linearly independent derivations $l_1, \cdots, l_r$, where $r \leq d$. Then the mapping of L onto the space $(l_1, \cdots, l_r)$ has Z for its kernel. Therefore, $\dim L = \dim Z + r$. But $r \leqq d$, and the lemma follows.

LEMMA 12. *If $A \lhd A_1 \lhd \lhd L$, if the centralizer of A in A_1 is zero, and if the centralizer of A_1 in L is zero, then the centralizer of A in L is zero.*

Proof. We will suppose on the contrary that A has centralizer $Z \neq (0)$ in L. Then $Z \cap A_1 = (0)$ by hypothesis. Consequently $P = (A_1, Z) \neq A_1$. Now $A \lhd A_1$ and $[A, Z] = (0)$. Therefore $A \lhd P$ and hence $Z \lhd P$. It follows that $P = A_1 + Z$. But $A_1 \lhd \lhd L$. Therefore $A_1 \lhd \lhd P \neq A$, and $A_1 \lhd A_2 \lhd \lhd P$, $A_2 \neq A_1$. Choose $a_2 \in A_2$, $a_2 \notin A_1$. By the form of P, $a_2 = a_1 + z$, $a_1 \in A_1$, $z \in Z$, where $z \neq 0$ since $a_2 \notin A_1$. Then $[z, A_1] \subseteq A_1 \cap Z$

since A_1 is an ideal of A_2 and Z an ideal of P. But $A_1 \cap Z = (0)$ and therefore $z \neq 0$ is in the centralizer of A_1, contradicting the hypothesis that A_1 had centralizer zero in L. Thus the lemma is established.

The proof of Theorem 17 will now be given. By Lemma 10, $L \cong D_0 \lhd D_1 \lhd D_2 \lhd \cdots \lhd D_n \lhd D_{n+1} \lhd \cdots$ where, for all l, D_{l-1} has zero centralizer in D_l. Suppose that D_{n-p} has zero centralizer in D_n. Then it follows from Lemma 12 that $D_{n-(p+1)}$ has zero centralizer in D_n. Since the statement is true for $p = 1$ by Lemma 10, it is true for all p; in particular, for $p = n$, whence $D_{n-n} = D_0 \cong L$ has zero centralizer in D_n. It follows from Theorem 16 that the centralizer of L^ω in D_n is contained in L^ω and hence has dimension c. Furthermore, L^ω is an ideal of D_n and hence by Lemma 11, $\dim D_n \leq d + c$, as the theorem asserts.

Since D_n has center zero for all n, the dimension of D_n is non-decreasing as a function of n, and it is immediate from the theorem that for some n all the derivations of D_n are inner. This result was announced by C. Chevalley [2] for Lie algebras of characteristic zero.

As a corollary of the Tower Theorem we have the following result.

THEOREM 18. *Let L be a Lie algebra with zero center over an arbitrary field. If L is isomorphic to an ideal K of the holomorph of L^ω such that K contains L^ω, then* $\dim D(L) = d + c$, *where again* $d = \dim D(L^\omega)$ *and c is the dimension of the center of L^ω. It follows that $D(L)$ has only inner derivations.*

Proof. We recall that the holomorph of L^ω is the semi-direct sum $H = L^\omega + D(L^\omega)$, where for $l \in L^\omega$, $\tilde{d} \in D(L^\omega)$, $[l, \tilde{d}]$ is defined to be $l\tilde{d}$. It is then obvious that L^ω is an ideal of H and hence of K. We have thus that $L^\omega \underline{\lhd} K \underline{\lhd} H$.

Now there are d linearly independent elements of H which determine linearly independent derivations of L^ω and hence of K. Furthermore, there are c linearly independent elements of H which are in the center of L^ω and hence are in K. But K is isomorphic to L and hence has no center. Therefore these c elements determine linearly independent derivations of K which are independent of the d elements above since they map L^ω into zero. Hence $\dim D(K) \geq d + c$ and consequently $\dim D(L) \geq d + c$. But from the Tower Theorem we know $\dim D(L) \leqq d + c$, and hence $\dim D(L) = d + c$, as asserted. Again it follows from the Tower Theorem that all the derivations of $D(L)$ are inner, and the proof is complete.

For the particular case where $L = L'$, the above result was proved by Zassenhaus.[10]

The following special case of the preceding theorem may also be of interest.

COROLLARY. *All the derivations of the holomorph of an Abelian Lie algebra are inner.*

Proof. Let A be an Abelian Lie algebra. Let $H = A + D(A)$ be the holomorph of A. Consider the subalgebra $L = A + (\tilde{j})$, where $\tilde{j}$ is the identity derivation of A; that is, $a\tilde{j} = a$ for all $a \in A$. Then L has no center. Furthermore it is easy to verify that $[\tilde{j}, \tilde{d}] = 0$ for all $\tilde{d} \in D(A)$, and hence L is an ideal of H. Finally, we note that $L^{\omega} = A$ and hence the previous theorem applies. Therefore $\dim D(L) = d + c$, where $d = \dim D(A)$, and $c = \dim A$ since A is Abelian. But $\dim L = d + c$, and since the centralizer of L in H is (0), it follows that $D(L) \cong H$, and the corollary follows.

LOUISIANA STATE UNIVERSITY.

BIBLIOGRAPHY.

[1] E. Cartan, *Sur la structure des groupes de transformation, finis et continus*, Paris, Libraire Vuibert (1933).

[2] C. Chevalley, "On groups of automorphism of Lie groups," *Proceedings of the National Academy of Sciences*, vol. 30 (1944), pp. 274-275.

[3] G. Hochschild, "Lie algebras and differentiations in rings of power series," *American Journal of Mathematics*, vol. 72 (1950), pp. 58-80.

[4] N. Jacobson, "Rational methods in the theory of Lie algebras," *Annals of Mathematics*, vol. 36 (1935), pp. 875-881.

[5] N. Jacobson, *The theory of rings*, American Mathematical Society, New York (1943).

[6] H. Weyl, "Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare Transformationen. II," *Mathematische Zeitschrift*, vol. 24 (1926), 328-376.

[7] H. Wielandt, "Eine Verallgemeinerung der invarianten Untergruppen," *Mathematische Zeitschrift*, vol. 45 (1939), pp. 209-244.

[8] H. Zassenhaus, "Über liesche Ringe mit Primzahlcharakteristik," *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 13 (1940), pp. 1-100.

[10] Zassenhaus [8], p. 52.

HOMOMORPHISMS OF A SEMIGROUP ONTO A GROUP.*

By R. R. Stoll.

Among the homomorphic images of a semigroup S (= a set closed with respect to an associative binary operation) there is at least one group, namely the unit group I. The question arises as to the existence of maximal group images of some sort for a given S. The answer is trivial if S contains a zero element, i. e. an element 0 such that $0x = x0 = 0$ for all $x \in S$, since then I is the only group image. In such cases the question has a broader scope if asked in connection with images of the form, a group with a zero element adjoined. These images exist if and only if S contain a prime ideal, according to Theorem 1.

In the present paper the notions of a maximal group image and maximal group with zero image are formulated for a semigroup S and the construction of such images is discussed in terms of "normal" subsystems (generalizations of a normal subgroup in the group case) of S, using a method devised by Dubreil [2]. A set of independent defining conditions for these normal subsystems is obtained. Finally, several well known semigroups which have unique maximal images of the types under considerations are discussed.

The most interesting examples are two classes of semigroups (regular sets of partial transformations of a set, and completely simple semigroups without zero) studied by Rees in [7] and [5] respectively, and one (semigroups having zeroid elements) studied by Clifford and Miller in [1].

For the first class, the group image discussed by Rees is shown to be maximal. For the second class, the maximal group image can be described in terms of a homomorphic image of the basis group for the semigroup. For the third class, the group of zeroid elements is found to be the maximal group image.

1. Group with zero images of a semigroup. The symbol G^* will be used to denote the group G with a zero element adjoined. In order to characterize those semigroups which admit some G^* as image, a definition is needed. By an ideal of a semigroup S is meant a subset P such that $SP, PS \subseteq P$. If, moreover, $ab \in P$ implies that at least one factor is in P, it is a prime ideal.

* Received March 8, 1950; presented to the American Mathematical Society, September 7, 1948.

THEOREM 1. *A semigroup S admits a group with zero as image if and only if it contains a prime ideal.*

Proof. Given $\alpha: S \to G^*$, it is easily verified that $\alpha^{-1}(0)$, where 0 is the zero element of G^*, is a prime ideal. Conversely, if S contains the prime ideal P, S can be mapped onto I^*, where I is the unit group, as follows: $S - P \to I$, $P \to 0$.

In view of this result, we restrict our attention in this section to semigroups S with a prime ideal. A homomorphism of S onto G^* is characterized by a pair (G^*, α) where α is the mapping function. For these pairs we make

Definition 1. $(G_1^*, \alpha_1) = (G_2^*, \alpha_2)$ if and only if $P(\alpha_1) = P(\alpha_2)$ where $P(\alpha_i)$ is the partition of S defined by the homomorphism α_i.

Equality of the above pair implies that G_1^* and G_2^* are isomorphic. As a representative of a class of equal pairs we may take the group with zero formed by the blocks of the common partition together with the homomorphism of inclusion. Henceforth (G^*, α) will denote such a pair.

In order to determine the possible distinct pairs (G^*, α) accompanying a given S, we use Dubreil's results [2] which we now describe. His basic notion is that of the right quotient of a complex $H \subset S$ by an element $a \in S$:

$$(H:a) = \{x \in S \mid ax \in H\}.$$

The principal right equivalence $\mathcal{E}_H$ induced in S by H is defined by $a \equiv b \pmod{\mathcal{E}_H}$ if and only if $(H:a) = (H:b)$. The right residue R_H is the set of all $x \in S$ for which $(H:x)$ is empty; R_H is a class mod $\mathcal{E}_H$. The definitions of the principal left equivalence ${}_H\mathcal{E}$ and left residue ${}_HR$ are clear. In the event $\mathcal{E}_H = {}_H\mathcal{E}$ and $R_H = {}_HR$, i. e. H is symmetric, the common equivalence relation $\mathcal{E}$ is a congruence relation, hence defines a homomorphic image $S/\mathcal{E}$ of S where the correspondence is that of inclusion. Those H's for which simultaneously $S/\mathcal{E}$ is a group with zero and with which any given homomorphism (G^*, α) can be duplicated, Dubreil calls the pseudo-normal and unitary subsemigroups N^* of S. His results in this connection are summarized in

THEOREM 2. *The correspondence $(G^*, \alpha) \to N^* = \alpha^{-1}(e)$, where e is the unit element of G, is a one-one correspondence between the set $\mathfrak{S}$ of distinct elements (G^*, α) and the set $\mathfrak{N}$ of pseudo-normal and unitary subsemigroups N^* of S. Moreover, in the correspondence $\alpha: S \to G^*$, the residue R (a prime ideal) of N^* is $\alpha^{-1}(0)$ where 0 is the zero element of G^*.*

For the complexes N^* it is possible to deduce from Dubreil's paper the

following set of independent defining conditions: (i) N^* is a subsemigroup, (ii) $a, b \in S$, and $ab \in N^*$ imply $ba \in N^*$, (iii) $a, b \in S$ and a and ab in N^* imply $b \in N^*$, (iv) the residue R_{N^*} is non-empty. Since the conditions (i)-(iii) imply that N^* is symmetric it is permissible to speak of the residue in (iv).

THEOREM 3. *The condition* (ii) *and* (iii) *above on a subsemigroup N^* of S are equivalent to the condition*

(v) *$a, b, c \in S$ and abc and ac in N^* imply $b \in N^*$.*

Proof. Assume that (v) holds. Then $ab \in N^*$ implies $(ab)^2 = a(ba)b \in N^*$, implies $ba \in N^*$, so that (ii) holds. Next, $a, ab \in N^*$ imply $aba, a^2 \in N^*$, hence $b \in N^*$ and (iii) holds. Conversely, if (ii) and (iii) hold, $abc, ac \in N^*$ imply $bca, ca \in N^*$, hence $b \in N^*$.

Definition 2. In $\mathfrak{G}$, call $(G_1^*, \alpha_1) < (G_2^*, \alpha_2)$ if and only if the partition $P(\alpha_2)$ is a refinement of $P(\alpha_1)$.

THEOREM 4. *$(G_1^*, \alpha_1) < (G_2^*, \alpha_2)$ if and only if $N_1^* \supseteq N_2^*$ and $R_1 = R_2$, where R_i is the residue of N_i^*, which in turn is the correspondent of (G_i^*, α_i) in accordance with Theorem 2.*

Proof. If $(G_1^*, \alpha_1) < (G_2^*, \alpha_2)$ then $N_2^* \subseteq N_1^*$ and $R_2 \subseteq R_1$. But by the definition of the residue, $x \in R_1$ implies $x \in R_2$, so that $R_1 \subseteq R_2$ and equality follows.

Conversely, if $N_2^* \subseteq N_1^*$ and $R_1 = R_2$ we shall prove that $a \equiv b (\mathrm{mod}\, \mathcal{E}_{N_2^*})$ implies $a \equiv b (\mathrm{mod}\, \mathcal{E}_{N_1^*})$. By assumption this is true for elements in the residue, so suppose that $(N_2^* : a) \neq O$. If $c \in (N_2^* : a)$ then ac and bc lie in N_2^*, and hence in N_1^*. If now $ax \in N_1^*$, we use the foregoing conditions (i), (v), (iii) for pseudo-normal and unitary subsemigroups in turn, to conclude that since $ax \cdot bc = a(xb)c$ and ac lie in N, so does xb, hence bx. Thus $(N_1^* : a) \subseteq (N_1^* : b)$. The reverse inequality is established similarly and equality holds; this means $a \equiv b (\mathrm{mod}\, \mathcal{E}_{N_1^*})$.

The set $\mathfrak{N}$ is the sum of the disjoint subsets $\mathfrak{N}_R = \{N^* \in \mathfrak{N} \mid R$ is the residue of $N^*\}$. This decomposition of $\mathfrak{N}$ defines a decomposition of $\mathfrak{G}$ into disjoint classes $\mathfrak{G}_R$. If $\mathfrak{N}_R$ is partially ordered by set inclusion and $\mathfrak{G}_R$ is partially ordered by Definition 2, Theorem 4 implies

THEOREM 5. *The one-one correspondence $(G^*, \alpha) \to N^*$ of the partially ordered sets $\mathfrak{G}_R$ and $\mathfrak{N}_R$ is an anti-isomorphism.*

Definition 3. By a maximal image of S of the type, a group with zero

element adjoined, is meant a maximal element in $\mathfrak{G}$ when partially ordered by Definition 2.

In view of the disjunction of $\mathfrak{G}$ into the classes $\mathfrak{G}_R$ and Theorems 4 and 5, the problem of finding maximal (G^*, α)'s is that of finding minimal N^*'s in $\mathfrak{N}_R$ for each of the residues R, i. e. each of the prime ideals in S.

2. Group images of a semigroup. The preceding section can be modified to the case of group images of a semigroup S (with or without a zero element), by omitting *'s and the notion of residues. The complexes N which generate group images of S, Dubreil calls the normal and unitary subsemigroups of S. They are characterized by (i), (ii), (iii) and $\overline{\text{(iv)}}$ (the negation of (iv)), or alternatively by (i), $\overline{\text{(iv)}}$ and (v). The latter is a set of independent defining conditions which can be extxracted from the redundant set devised by Levi [3, 4] for defining his "complete and normal subsemigroups" in connection with the problem of determining all group images of a semigroup.

3. Examples of maximal group with zero images. We now obtain the maximal images of the form, a group with zero, for several semigroups.

Example 1. The semigroup M of all $n \times n$ matrices over a field. It is a routine matter to conclude that the only ideals of M, apart from M, are the sets M_r consisting of all matrices of rank $\leqq r$, $r = 0, 1, \cdots, n-1$ and that among these only M_{n-1} is a prime ideal. Associated with this unique residue is a unique minimal pseudo-normal and unitary subsemigroup, namely the unit matrix I. The corresponding maximal group with zero image is, of course, the group of non-singular matrices with zero adjoined. The only interest in this obvious result is the uniqueness of the maximal image.

Example 2. The semigroup of all correspondences of a finite set C onto subsets of C. The map of C onto the empty subset is the zero element. An analysis along the lines of Example 1 yields the group of al permutations of C together with zero as the unique maximal image. The same conclusion holds for

Example 3. The semigroup of all partial transformation of a finite set C. Rees [7] has defined a partial transformation $\alpha: x \to x\alpha$ of a set C as a one-one correspondence of a subset C onto a subset of C. If $A(\alpha)$ denotes the set of elements for which α is defined and $V(\alpha)$ the set of image elements, the product $\alpha\beta$ of two partial transformations is defined by iteration over $V(\alpha) \cap A(\beta)$. The zero element ζ is the partial transformation with both $V(\zeta)$ and $A(\zeta)$ the empty set; $\alpha\beta = \zeta$ if and only if $V(\alpha) \cap A(\beta)$ is empty.

4. Examples of maximal group images.

Example 4. A regular set of partial transformations of a set C. This is a motion introduced by Rees [7]. He assumes that the product of the transformations α, β is defined only if $V(\alpha) \cap A(\beta)$ is non-empty and calls a set Σ of partial transformations regular if (a) $\alpha, \beta \in \Sigma$ imply $\alpha\beta$ is defined and in Σ, and (b) $\alpha \in \Sigma$ implies $\alpha^{-1} \in \Sigma$, where α^{-1} is the transformation $x\alpha \to x$. A regular set Σ is a semigroup which may or may not have a zero element. Rees obtains a group image for Σ as follows. If for $\alpha, \alpha' \in \Sigma$, $A(\alpha') \subseteq A(\alpha)$ and for all $x \in A(\alpha')$, $x\alpha' = x\alpha$, α' is a subtransformation of α. The relation $\mathfrak{R} : \alpha \equiv \beta (\mathrm{mod}\ \mathfrak{R})$ if and only if α and β have a common subtransformation in Σ is a congruence relation over Σ and the quotient $\Sigma/\mathfrak{R}$ of Σ by $\mathfrak{R}$ is a group homomorphic to Σ.

We now show that $\Sigma/\mathfrak{R}$ is the unique maximal group image of Σ. If $\alpha \in \Sigma$, so is the identity transformation $\epsilon = \alpha\alpha^{-1}$. Thus the unit residue class U mod $\mathfrak{R}$ consists of all elements having a subtransformation in common with any one fixed identity transformation. If on the other hand N is any normal and unitary subsemigroup in Σ, it satisfies (i), (ii), (iii), (iv) and (v) of § 1. From these conditions we conclude:

(1) If $\alpha \in N$, so is α^{-1}; thus N contains an identity $\epsilon = \alpha\alpha^{-1}$. For along with $\alpha = \alpha\alpha^{-1}\alpha$, N contains $\alpha\alpha$ by (i), hence α^{-1} by (v).

(2) If β has $\gamma \in N$ as a subtransformation then $\beta \in N$. For $\gamma\gamma^{-1}$, $(\gamma\gamma^{-1})\beta = \gamma \in N$ imply $\beta \in N$ by (iii).

(3) If β and the identity ϵ have a common subtransformation $\gamma \in \Sigma$ then $\beta \in N$. For γ, as a subtransformation of ϵ is idempotent and hence in N. The assertion then follows from (2).

Thus every normal and unitary N contains the above U and the uniqueness of U as the minimal normal and unitary subsemigroup is established.

Example 5. A completely simple semigroup without zero. Rees [5, 6] has defined such a semigroup S by the conditions, (a) the only ideal in S is S itself, (b) S contains a primitive idempotent, i. e. an idempotent e such that there is no idempotent $f \neq e$ for which $ef = fe = f$. Such a semigroup can be described as the set of all triplets

$$(a; \alpha, \beta) \qquad a \in H, \qquad \alpha \in A, \qquad \beta \in B,$$

where H is a group, the basis group for S, and A and B are two sets of indices. To each pair α, β of indices ($\alpha \in A, \beta \in B$) there is assigned some fixed element $p_{\beta\alpha} \in H$ and multiplication is defined by $(a; \alpha, \beta)(b; \alpha', \beta')$

$= (ap_{\beta\alpha}b; \alpha, \beta')$. We shall call the $p_{\beta\alpha}$ the constants of S. It is possible to alter these constants somewhat without changing S. For example, if 1 is taken to be a member of both sets A, B of indices, we may choose $p_{\alpha 1} = p_{1\beta} = e$, the unit element in H, for all $\alpha \in A, \beta \in B$. We assume this is done. The idempotents of S are the elements $e_{\alpha\beta} = (p^{-1}{}_{\beta\alpha}; \alpha, \beta)$. The set of elements $(a; \alpha, \beta)$ where a varies over H and α, β are fixed, form a group isomorphic to H, so that S is the sum of disjoint groups, each isomorphic to H.

THEOREM 6. *Let S denote a completely simple semigroup without zero, with basis group H, index sets A, B and constants $p_{\beta\alpha}$. Let P denote the (unique) minimal normal subgroup of H containing the $p_{\beta\alpha}$. Then S contains a unique minimal normal and unitary subsemigroup* $N = \sum_{\alpha\in A, \beta\in B} (P; \alpha, \beta)$, *where $(P; \alpha, \beta)$ is the totality of elements $(p; \alpha, \beta)$ with $p \in P$. Thus S has a unique maximal group image. This group is isomorphic to H/P.*

Proof. We first show that a normal and unitary subsemigroup M of S must contain N, using the conditions in § 1. Since M must contain the idempotents $e_{\alpha\beta}$ of S, it contains $e_{1\beta}e_{\alpha 1} = (p_{\beta\alpha}; 1, 1)$. Moreover $(p_{\beta\alpha}; 1, 1)(p_{\beta\alpha}{}^{-1}; 1, 1) = e_{11} \in M$, hence by (iii) $(p_{\beta\alpha}{}^{-1}; 1, 1) \in M$. The equations

$$(x^{-1}; 1, 1)(x; 1, 1) = e_{11} \in M, \quad (x^{-1}; 1, 1)(xp_{\beta\alpha}x^{-1}; 1, 1)(x; 1, 1) \in M$$

and (v) imply that $(xp_{\beta\alpha}x^{-1}; 1, 1) \in M$; similarly for $(xp_{\beta\alpha}{}^{-1}x^{-1}; 1, 1)$. Since M is closed, it follows that $(P; 1, 1) \subset M$. Finally, the equations

$$e_{11}e_{11} = e_{11} \in M, e_{11}\ (p; \alpha, \beta)e_{11} = (p; 1, 1) \in M \text{ for } p \in P,$$

and (v) imply that $(p; \alpha, \beta) \in M$. Thus $M \supset N$.

To show that N is normal and unitary it suffices to verify that (i), (iv) and (v) hold. Clearly N is closed. As for (iv), multiplication on the right of an element $(a; \alpha, \beta) \in S$ by $((p_{\alpha\delta}ap_{\beta\gamma})^{-1}; \gamma, \delta)$ gives $e_{\alpha\delta} \in N$. Finally, suppose that the products $(a; \alpha, \beta)(b; \gamma, \delta)(c; \lambda, \mu)$ and $(a; \alpha, \beta)(c; \lambda, \mu)$ are in N. These products then have the form $(p_1; \alpha, \mu)$ and $(p_2; \alpha, \mu)$ respectively, where $p_i \in P$. If one calculates $a^{-1}p_1p_2{}^{-1}a$ it is immediately seen that $b \in P$, hence that $(b; \gamma, \delta) \in N$. Thus (v) holds.

The proof of the isomorphism of $S/\mathcal{E}_N$ and H/P is immediate. One finds that the cosets of P in H describe the equivalence classes in $S \bmod \mathcal{E}_N$, i. e. an equivalence class in $S \bmod \mathcal{E}_N$ has the form $\sum_{\alpha\in A, \beta\in B} (hP; \alpha, \beta)$. The correspondence of this class with the coset hP of P in H establishes the isomorphism.

Example 6. Semigroups having zeroid elements. Clifford and Miller [1] have defined an element u of a semigroup S as a zeroid element if, for each element a of S, there exist x and y in S such that $ax = ya = u$. If the set U of all zeroid elements in S is non-empty, U is a group and a homomorph of S under the correspondence $\zeta: x \to xz$ where z is the unit element of U. The set $J = \zeta^{-1}(z)$ has been characterized by Clifford and Miller as the set of all elements in S for which z is a zero element.

We shall prove that U is the unique maximal group image of S by verifying that if N is any normal and unitary subsemigroup in S then $N \supseteq J$. By (iv), for an element $a \in J$, there exists an $x \in S$ such that $ax \in N$. As an idempotent, $z \in N$, hence $zax = zx \in N$ by (i). But z and zx in N imply $x \in N$ by (iii). Finally, x and $ax \in N$ imply $a \in N$ by (iii). Thus $J \subseteq N$. Moreover the map $S \to S/\mathcal{E}_J$ (where the homomorphism is that of inclusion) is the map $\zeta: x \to xz$.

5. Remarks. In contrast to the foregoing examples, the additive semigroup P of positive integers is an instance of a semigroup having no maximal group images. For clearly the group images of P are precisely the totality of finite multiplicative cyclic groups and among these there is no maximal group.

The problem of finding useful reformulations of the condition obtained earlier (namely the existence of a minimal normal and unitary subsemigroup) for the existence of a maximal group image appears to be difficult. The same is true for the problem of determining when a given semigroup has a unique maximal group image. These matters will be discussed elsewhere.

LEHIGH UNIVERSITY.

REFERENCES.

[1] A. H. Clifford and D. D. Miller, "Semigroups having zeroid elements," *American Journal of Mathematics*, vol. 70 (1948), pp. 117-125.

[2] Paul Dubreil, "Contribution à la théorie des demi-groupes," *Mémoires de l'Académie des Sciences*, ser. 2, vol. 63 (1941), pp. 1-52.

[3] F. W. Levi, "On semigroups," *Bulletin of the Calcutta Mathematical Society*, vol. 36 (1944), pp. 141-146.

[4] ———, "On semigroups II," *ibid.*, vol. 38 (1946), pp. 123-124.

[5] D. Rees, "On semigroups," *Proceedings of the Cambridge Philosophical Society*, vol. 36 (1940), pp. 387-400.

[6] ———, "Note on semigroups," *ibid.*, vol. 37 (1941), pp. 434-435.

[7] ———, "On the group of a set of partial transformations," *Journal of the London Mathematical Society*, vol. 22 (1947), pp. 281-284.

A NOTE ON A FUNCTIONAL EQUATION.*

By H. P. THIELMAN.

We consider the functional equation

$$\phi[F(x,y)] = \phi(x) + \phi(y) \tag{1}$$

and determine the conditions on a polynomial $F(x,y)$ in order that (1) can have a continuous strictly monotone (say increasing) solution. If $F(x,y)$ is $x+y$ or xy, we have the classical cases of Cauchy's functional equations. The cases when $F(x,y)$ is a general linear function,[1] and also the case when it is $x+y+nxy$, have been treated.[2] In this paper the case when $F(x,y)$ is a polynomial of degree greater than unity is settled, as follows:

THEOREM. *If $F(x,y)$ is a polynomial of degree greater than unity for which* (1) *has a continuous monotone increasing solution, then $F(x,y)$ must be of the form*

$$F(x,y) = \{(ax+b)(ay+b) - b\}/a, \qquad a > 0.$$

Proof. By hypothesis, equation (1) has a strictly monotone increasing, continuous solution $\phi(x)$. Hence the inverse $\phi^{-1}(x)$ will exist and we can write (1) as

$$F(x,y) = \phi^{-1}[\phi(x) + \phi(y)]. \tag{2}$$

In order to simplify the notation we shall denote $F(x,y)$ by $x \circ y$. This can be interpreted as an operation which, because of (2), has the following properties: (i) $x \circ y = y \circ x$, commutativity; (ii) $x \circ (y \circ z) = (x \circ y) \circ z$, associativity; (iii) $x' \circ y > x \circ y$ if $x' > x$, strict monotony; (iv) continuity. [From (iii) and (i) it follows that $x \circ y' > x \circ y$ if $y' > y$, $x \circ x < x \circ y < y \circ y$ if $x < y$. Aczél has shown that any operation satisfying the above conditions can be written in form (2)].

Let $x \circ y$ be a polynomial of degree n in x, and m in y. Then, because of (i), $n = m$. Because of (ii), $(x \circ y) \circ z = x \circ (y \circ z)$. The left hand side of this equation is a polynomial of degree n in z, the right hand side a

* Received July 8, 1949.

[1] J. Aczél, *Comm. Math. Helvetici*, vol. 21 (1948), pp. 247-252.

[2] H. P. Thielman, *American Mathematical Monthly*, vol. 56 (1949), pp. 452-457.

polynomial of degree n^2 in z. Hence $n = 1$. Thus $x \circ y$ is a symmetric polynomial of degree 2 in x and y combined. It can be written as $x \circ y = Bxy + C(x + y) + D$.

Let us denote $x \circ x$, which is of degree 2, by $P_2(x)$, and $x \circ (x \circ x)$, which is of degree 3, by $P_3(x)$. By (ii), $[x \circ x] \circ [x \circ (x \circ x)] = [x \circ (x \circ x)] \circ (x \circ x)$, which states that $P_2(x)$ and $P_3(x)$ are commutative. But it is easily proved, by equating coefficients, that if a quadratic polynomial commutes with a cubic one, then it must be of the form $\lambda^{-1}[\lambda^2(x)]$ or $\lambda^{-1}[\cos 2 \arccos \lambda(x)]$, where $\lambda(x)$ is linear, say $ax + b$, and $\lambda^{-1}(x)$ is its inverse. Thus either

a) $x \circ x = Bx^2 + 2Cx + D \equiv \{(ax + b)^2 - b\}/a$

or

b) $x \circ x = Bx^2 + 2Cx + D \equiv \{2(ax + b)^2 - 1 - b\}/a$.

We first consider the identity a). It implies that $B = a$, $C = b$, $D = (b^2 - b)/a$, and $F(x, y) = \{(ax + b)(ay + b) - b\}/a$. It is easily seen that conditions (i), (ii), (iii), (iv) are satisfied if $x > -b/a$, $y > -b/a$, $a > 0$. If we let $\phi(x) = \log(ax + b)$, then $F(x, y) = \phi^{-1}[\phi(x) + \phi(y)]$, or

$$\phi[\{(ax + b)(ay + b) - b\}/a = \phi(x) + \phi(y).$$

[By a method similar to the one used *loc. cit.*,[2] it can be shown that every continuous solution of this last functional equation is of the form $k \log(ax + b)$, where k is an arbitrary constant.]

We next consider the identity b). It can be written as $Bx^2 + 2Cx + D \equiv 2ax^2 + 4bx + (2b^2 - 1 - b)/a$, from which it follows that

$$x \circ y = F(x, y) = 2axy + 2b(x + y) + (2b^2 - 1 - b)/a.$$

Condition (ii) is not satisfied for $(x \circ x) \circ z \neq x \circ (x \circ z)$, as can be easily verified. Hence this case is eliminated and the theorem is proved.

The theorem could be stated in the terminology of the literature[3] dealing with operations as follows: Every rational, integral, associative and commutative operation which is of degree greater than unity is of the form $\lambda^{-1}[\lambda(x)\lambda(y)]$, where $\lambda(x)$ is a linear function.

COROLLARY. *Every associative and commutative operation $x \circ y$ which for $x = y$ reduces to a polynomial of degree greater than unity must be such that either $x \circ x = \lambda^{-1}[\lambda^2(x)]$ or $x \circ x = \lambda^{-1}[\cos 2 \arccos \lambda(x)]$.*

The proof of this corollary was given in the process of proving the theorem.

---

[3] J. Aczél, *Bulletin of the American Mathematical Society*, vol. 54 (1948), pp. 392-400.

That there exist algebraic associative, commutative and increasing operations which are not integral is shown by the example

$$x \circ y = \{(ax+b)(ay+b) - [1-(ax+b)^2]^{\frac{1}{2}}[1-(ay+b)^2]^{\frac{1}{2}} - b\}/a,$$

which can be written as

$$x \circ y = \{\cos[\arccos(ax+b) + \arccos(ay+b)] - b\}/a.$$

In the interval $-b/a \leqq x, y \leqq (1-b)/a$, $(a > 0)$, this function satisfies all the conditions (i), (ii), (iii), (iv), and it also illustrates the corollary. Obviously this last function can serve in the functional equation (1), whose continuous solutions then are all of the form $\phi(x) = k \arccos(ax+b)$, where k is any constant. But since $x \circ y$ in this case is not a polynomial we shall not concern ourselves with it in this paper. It might be of interest to note that every associative and commutative operation $x \circ y$ which for $x = y$ reduces to a polynomial of degree greater than unity will generate an "entire set of commutative polynomials" [4] which must be either linear transforms of powers of x or of the trigonometric polynomials $\cos n \arccos x$.[5] By a linear transform of a function $f(x)$ is meant the function $\lambda^{-1}[f(\lambda(x))]$, where $\lambda(x)$ is any linear function $ax+b$, $a \neq 0$. To generate these polynomials we need to iterate $x \circ x$ with itself. It should be mentioned that if we had replaced "monotone increasing" by "monotone decreasing," the results would still be true but for intervals different from those indicated above.

IOWA STATE COLLEGE.

[4] B. Vinograde, *American Mathematical Monthly*, vol. 56 (1949), pp. 377-379.

[5] This follows from the results of J. F. Ritt, *Transactions of the American Mathematical Society*, vol. 25 (1923), pp. 339-448.